MME 2017

35th International Conference **Mathematical Methods in Economics**

September 13th - 15th, 2017, Hradec Králové, Czech Republic

Conference Proceedings

35th International Conference

Mathematical Methods in Economics

MME 2017

Conference Proceedings

Hradec Králové, Czech Republic
September 13th – 15th, 2017

ISBN 978-80-7435-678-0

Faculty of Informatics and Management, University of Hradec Králové

Programme Committee

doc. RNDr. Helena Brožová, CSc.
doc. Mgr. Ing. Martin Dlouhý, Dr. MSc.
Mgr. Jan Draesslar, Ph.D.
doc. Ing. Jan Fábry, Ph.D.
prof. RNDr. Ing. Petr Fiala, CSc., MBA
doc. Ing. Jana Hančlová, CSc.
prof. Ing. Josef Jablonský, CSc.
doc. RNDr. Ing. Miloš Kopa, Ph.D.
doc. RNDr. Ing. Ladislav Lukáš, CSc.
prof. RNDr. Jan Pelikán, CSc.
doc. Dr. Ing. Miroslav Plevný
doc. RNDr. Petra Poulová, Ph.D.
doc. RNDr. Pavel Pražák, Ph.D.
prof. RNDr. Jaroslav Ramík, CSc.
Ing. Karel Sladký, CSc.
prof. RNDr. Hana Skalská, CSc.
doc. Ing. Tomáš Šubrt, Ph.D.
doc. RNDr. Jana Talašová, CSc.
prof. RNDr. Milan Vlach, DrSc.
prof. RNDr. Karel Zimmermann, DrSc.
prof. Ing. Miroslav Žižka, Ph.D.

Editor: Pavel Pražák

Contents

A DEA approach for selecting performance measures in presence of negative data

Maryam Allahyar[1], Mehdi Toloo[2]

Abstract. One of the basic assumptions in developing the classical Data envelopment analysis (DEA) models is that all input and output measures are non-negative, whereas we may encounter some cases including negative data in real word applications. On the other hand, there is an empirical rule of thumb that if the total number of input and output measures is high in comparison with the number of units, then a large percentage of units will be identified as efficient which is an unreasonable result. Considering this two issues in DEA, this current paper develops a selecting method which chooses important measures with the aim of keeping the rule of thumb in the presence of negative data.

Keywords: Data Envelopment Analysis, Negative data, Directional Distance Function, Selective measures.

JEL Classification: C61, C67
AMS Classification: 90C11

1 Introduction

Data envelopment analysis (DEA) is a non-parametric approach based on mathematical programming to evaluate a set of peer decision making units (DMUs) with multiple inputs and multiple outputs. The CCR [1] and BCC [2] models are two basic DEA models. Inputs and outputs play a key role in DEA and some approaches have been developed for classifying inputs and outputs in DEA (for more details see Toloo [3, 4]). However, the classical DEA models are formulated with the assumption that all input and output measures are non-negative, whereas this assumption is very restrictive in many real applications. The DEA research on facing with negative data can be mainly divided into three groups. The first one is that the negative inputs are treated as positive outputs and the negative outputs are treated as positive inputs (Scheel [5], Zhu [6]). The second one is the approach applying DEA models with translation invariance to make all negative data positive by adding a big enough scalar to the negative variables (such as Charnes et al. [7]; Seiford & Zhu [8]). The last one is handling negative variables without translation of the original data, e.g. Portela et al. [9], Sharp et al. [10], Emrouznejad et al. [11] and Cheng et al. [12]. In a recent paper, Allahyar and Rostamy [13] addressed the efficiency measure by developing the directional distance function and then suggested a method aims at discovering the state of returns to scale (RTS) in the presence of negative input and output values.

Another issue concerning DEA models is the relationship between the number of performance measures and the number of DMUs. Empirically, if the number of DMUs is low in comparison with the combined number of performance measures (inputs and outputs), then a large percentage of DMUs will be identified as efficient which is an unreasonable result. To have a reliable result, there is a rough rule of thumb that expresses the relationship between the number of measures and the number of DMUs. Toloo et al. [14] and Toloo & Tichý [15] proposed two selecting models under constant returns to scale (CRS) and variable returns to scale (VRS) assumptions, respectively, to meet the rule of thumb.

To deal with these two challenging issues, simultaneously, the current paper proposes a selecting DEA model, which chooses some measures with the aim of keeping the rule of thumb in the presence of negative data.

The paper is organized into five sections. Section 2 briefly reviews the directional distance function and efficiency evaluation in the presence of negative data proposed by Allahyar and Rostamy [13] as well as the

[1] Young Researchers and Elite Club, Yadegar-e-Imam Khomeini (RAH) Shahre Rey Branch, Islamic Azad University, Tehran, Iran, allahyar_maryam@yahoo.com

[2] Department of Systems Engineering, Technical University of Ostrava, Sokolska tř. 33, Ostrava 1, Ostrava, Czech Republic, mehdi.toloo@vsb.cz

selecting model of [15]. The suggested selecting approach with negative data is explained in Section 3. The penultimate section illustrates the applicability of the suggested model. Finally, we have come to conclusion in Section 5.

2 The relative efficiency measure of negative data

In this section, we give a brief review of some preliminary concepts about the directional distance function and efficiency evaluation in the presence of negative data.

Consider a set of n DMUs, DMU_j ; $j \in \{1, \dots, n\}$ and let x_{ij}, $i \in \{1, \dots, m\}$, be m different semi-positive inputs and y_{rj}, $r \in \{1, \dots, s\}$ be s different semi-positive outputs for DMU_j. The generic directional distance model (DD model hereafter) as the most general non-oriented model developed by Chambers et al. ([16],[17]) under CRS technology is presented as:

$$
\begin{aligned}
&\mathrm{z}_1 = \max\ \theta \\
&\text{s.t.} \\
&\textstyle\sum_{j=1}^{n} \lambda_j\, x_{ij} \le x_{io} - \theta g_{xi} \qquad i = 1, \dots, m \\
&\textstyle\sum_{j=1}^{n} \lambda_j\, y_{rj} \ge y_{ro} + \theta g_{yr} \qquad r = 1, \dots, s \\
&\lambda_j \ge 0 \qquad\qquad\qquad\qquad j = 1, \dots, n
\end{aligned}
\tag{1}
$$

where the nonzero *directional* vector $(\boldsymbol{g}_x, \boldsymbol{g}_y) = (g_{x1}, \dots, g_{xm}, g_{y1}, \dots, g_{ys})$ is adopted aim at input contraction and output expansion, simultaneously. Since data are positive a usual choice for the direction vectors $(\boldsymbol{g}_x, \boldsymbol{g}_y)$ is the observed input and output levels of DMU_o, i.e. $(\boldsymbol{g}_x, \boldsymbol{g}_y) = (\boldsymbol{x}_o, \boldsymbol{y}_o)$.

Now, assume that the performance measures are allowed to take negative as well as positive values. Allahyar and Rostamy [13] with the aim of extending the DD model (1) in the presence of negative data, formulated the following Linear Programming (LP) model for assessing DMU_o , $o \in \{1, \dots, n\}$ under variable returns to scale (VRS):

$$
\begin{aligned}
&z_2 = \max \theta \\
&\text{s.t.} \\
&\textstyle\sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io} - \theta |x_{io}| \qquad \forall i \\
&\textstyle\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro} + \theta |y_{ro}| \qquad \forall r \\
&\textstyle\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \ge 0 \qquad\qquad\qquad \forall j
\end{aligned}
\tag{2}
$$

In the model above, actually the direction vector $(-|\boldsymbol{x}_o|, |\boldsymbol{y}_o|) = (-|x_{1o}|, \dots, -|x_{mo}|, |y_{1o}|, \dots, |y_{so}|)$ is considered by using the absolute values of data. From now on, "*" indicates the optimal values. A trivial verification shows that $\theta^* \ge 0$. DMU_o is efficient if $\theta^* = 0$; otherwise DMU_o is inefficient. Indeed the value of θ^* signifies the measure of inefficiency.

2.1 DEA selecting model

As it is mentioned in the previous section, empirically, there is a rough rule of thumb in the envelopment model expressing the relationship between the number of DMUs (n) and the combined number of inputs and outputs ($m + s$). Specifically, this rule says that if $n < \max\{3(m + s), m \times s\}$, then most DMUs are classified as efficient through the assessing model which is not acceptable. In this case, in order to have a sharper discrimination among DMUs, it is needed to adopt a selecting approach such that it selects measures optimally in a way that the number of selected measure and DMUs meet the rule of thumb. In order to cope with these situations, Toloo & Tichý [15] extended the envelopment BCC model and presented the following selecting mixed integer programing (MIP) model:

$$
\begin{aligned}
&\mathrm{z}_3 = \min \theta - \varepsilon\left(\sum_{i=1}^{m} d_i^x s_i^x + \sum_{r=1}^{s} d_r^y s_r^y\right) \\
&\text{s. t.} \\
&\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^x = \theta x_{io} + M(1 - d_i^x) && \forall i \\
&\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^y = y_{ro} - M\left(1 - d_r^y\right) && \forall r \\
&\sum_{r=1}^{s} d_r^y + \sum_{i=1}^{m} d_i^x \le \min\left\{\left[\frac{n}{3}\right], 2\sqrt{n}\right\} \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&d_i^x, d_r^y \in \{0,1\} && \forall i, \forall r \\
&s_i^x, s_r^y \ge 0 && \forall i, \forall r \\
&\lambda_j \ge 0 && \forall j
\end{aligned}
\tag{3}
$$

where $\varepsilon > 0$ is a small number and M is a large positive number. Moreover, auxiliary binary variables d_i^x and d_r^y are introduced for each selective input and output measure, respectively. The i^{th} selective input (r^{th} selective output) is selected if $d_i^x = 1\left(d_r^y = 1\right)$, otherwise, i.e. $d_i^x = 0\ \left(d_r^y = 0\right)$, the i^{th} selective input (r^{th} selective output) will not selected. Toloo & Tichý [15] proved that the rough rule of thumb $n \ge \max\{3(m+s), m \times s\}$ translates into the constraint $\sum_{r=1}^{s} d_r^y + \sum_{i=1}^{m} d_i^x \le \min\{[n/3], 2\sqrt{n}\}$, mathematically. It is worth mentioning that Toloo & Tichý [15] also suggested to add constraints $\sum_{i=1}^{m} d_i^x = p$ and $\sum_{r=1}^{s} d_r^y = q$ to model (3) when the manager is interested in selecting exactly p out of selective inputs and q out of selective outputs.

3 The proposed selecting DD model with negative data

Inspired by the selecting model (3), this section extends model (2) and develops a new selecting DD approach, which permits the performance measures to take both positive and negative values. To do this, we formulate the following MIP model:

$$
\begin{aligned}
&\mathrm{z}_4 = \max \beta \\
&\text{s. t.} \\
&\sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io} - \beta|x_{io}| + M(1 - b_i^x) && \forall i \\
&\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro} + \beta|y_{ro}| - M\left(1 - b_r^y\right) && \forall r \\
&\sum_{r=1}^{s} b_r^y + \sum_{i=1}^{m} b_i^x \le \min\left\{\left[\frac{n}{3}\right], 2\sqrt{n}\right\} \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&b_i^x, b_r^y \in \{0,1\} && \forall i, \forall r \\
&\lambda_j \ge 0 && \forall j
\end{aligned}
\tag{4}
$$

where analogous to model (3), for each selective input and output measure we define the auxiliary binary variables b_i^x and b_r^y, respectively. To clarify the model formulation, some explanations are provided as below:

If $b_i^x = 0$, then the constraint $\sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io} - \beta|x_{io}| + M(1 - b_i^x)$ implies $\sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io} - \beta|x_{io}| + M$ which is not binding; otherwise, i.e. $b_i^x = 1$, this constraint is weakened to the active constraint $\sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io} - \beta|x_{io}|$.

Likewise,

$$
\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro} + \beta|y_{ro}| - M\left(1 - b_r^y\right) \Rightarrow \begin{cases} \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro} + \beta|y_{ro}| - M \text{ (nonbinding)}, & \text{if } b_r^y = 0 \\ \sum_{j \in J} \lambda_j y_{rj} \ge y_{ro} + \beta|y_{ro}| \quad \text{(active)}, & \text{if } b_r^y = 1. \end{cases}
$$

The following theorem shows that the proposed model (4) is succeed in decreasing the percentage of efficient DMUs.

Theorem 1. $z_2^* \le z_4^*$

Proof. Suppose that model (2) is solved and the optimal solution $(\beta^*, \boldsymbol{\lambda}^*) \in \mathbb{R}^{n+1}$ is at hand. An easy computation shows that $(\beta^*, \boldsymbol{\lambda}^*, \boldsymbol{b}^x, \boldsymbol{b}^y) \in \mathbb{R}^{n+\mathrm{m}+\mathrm{s}+1}$ is a feasible solution for model (4) with the objective function value $z_2^* = \beta^*$ when $\sum_{r=1}^{s} b_r^y + \sum_{i=1}^{m} b_i^x \le \min\{[n/3], 2\sqrt{n}\}$. Since model (4) is a max-type problem, we achieve $z_2^* \le z_4^*$, which completes the proof. ■

It should be mentioned here that model (4) has been formulated with the assumption that all inputs and outputs are selective measures. However, in some cases, the manager believes that some performance measures are

necessary for the system and we fix such performance measures and select from the remaining inputs and outputs as selective measures.

Now, after solving model (4) for each DMU and obtaining a set of optimal $\boldsymbol{b}^x$ and $\boldsymbol{b}^y$ in the proposed individual approaches, one criterion for selecting measures would then be to base it on the majority choice among the DMUs. The corresponding efficiency scores achieved for each DMU in the light of the final selected measures are achieved by applying the assessing model (**2**).

4 Application

In order to apply the proposed selecting models to a real data set, we use the data set of "the notional effluent processing system" whose data has been extracted from Sharp et al. [10] and includes 13 DMUs with one positive input (x_1: Cost), one negative input (x_2: Effluent), one positive output (Saleable output: y_1) and two negative outputs (Methane: y_2 and CO_2: y_3). Table 1 contains the input–output data.

Table 1 Input–output data

DMU	Inputs		Outputs		
	Cost	**Effluent**	**Saleable**	**Methane**	**CO_2**
1	1.03	−0.05	0.56	−0.09	−0.44
2	1.75	−0.17	0.74	−0.24	−0.31
3	1.44	−0.56	1.37	−0.35	−0.21
4	10.8	−0.22	5.61	−0.98	−3.79
5	1.3	−0.07	0.49	−1.08	−0.34
6	1.98	−0.1	1.61	−0.44	−0.34
7	0.97	−0.17	0.82	−0.08	−0.43
8	9.82	−2.32	5.61	−1.42	−1.94
9	1.59	0.00	0.52	0.00	−0.37
10	5.96	−0.15	2.14	−0.52	−0.18
11	1.29	−0.11	0.57	0	−0.24
12	2.38	−0.25	0.57	−0.67	−0.43
13	10.3	−0.16	9.56	−0.58	0.00

As can be seen, this data set with $n = 13$, $m = 2$, and $s = 3$ does not meet the rough rule of thumb, i.e. we have $13 < 15 = \max\{3(m + s), m \times s\}$. To overcome this deficiency, the proposed selecting model (4) should be applied to the data set in Table 1 with the aim of decreasing the number of measures. In this case, it is supposed that $p = 1$ and $q = 2$. Table 2 exhibits the optimal value of b^{x*} and b^{y*}.

Table 2 The optimal value of $\boldsymbol{b}^x$ and $\boldsymbol{b}^y$

DMU	b_1^{x*}	b_2^{x*}	b_1^{y*}	b_2^{y*}	b_3^{y*}	Acceptable Efficiency
1	0	1	1	0	1	1
2	0	1	1	0	1	0.763
3	0	1	1	1	0	0
4	1	0	0	1	1	0.63
5	0	1	1	0	1	1
6	0	1	1	0	1	0.94
7	0	1	1	0	1	0.81
8	1	0	0	1	1	0
9	0	1	1	0	1	1
10	0	1	1	0	1	0.71
11	0	1	1	0	1	0.89
12	0	1	1	1	0	0.68
13	1	0	0	1	1	0

The optimal solution implies that 'Effluent', 'Saleable output' and 'CO2' are determined as selected measures. In other words, the majority of DMUs have their best performance in light of the set of selected measures $\{x_2, y_1, y_3\}$. Last column of Table 2 demonstrates the acceptable efficiencies, which are obtained by applying model (2) to the selected data set. It can be extracted that there are 3 efficient DMUs $\{3, 8, 13\}$, while there are more efficient units in the light of all performance measures. Actually, these results show that the formulated selecting model in the presence of negative data successfully decreased the percentage of efficient DMUs.

5 Conclusion

This paper dealt with two challenging issues in DEA applications, the presence of negative measures in data set and holding the rule of thumb in modeling, which expresses the relationship between the total performance measures and the number of DMUs. In particular, inspired by the selecting method of Toloo & Tichý [15], we extended the assessing model proposed by Allahyar and Rostamy [13] aims at formulating a selecting model which identifies some important measures so that the selected data set satisfies the rule of thumb in the presence of negative data. The applicability of the suggested method is illustrated by a real data set of "the notional effluent processing system". The proposed approach in this study can be extended to imprecise data in DEA (see Toloo [18], Toloo and Ertay [19], and [20])

Acknowledgements

This paper was supported by SGS Project No. SP2017/141 VŠB-Technical University of Ostrava**.**

References

[1] Charnes, A., Cooper, W. W., and Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operation Research* **2**(6)(1978), 429–444.

[2] Banker, R.: Estimating most productive scale size using data envelopment analysis. *European Journal of Operation Research* **17**(1)(1984), 35–44.

[3] Toloo, M., Alternative solutions for classifying inputs and outputs in data envelopment analysis, *Computers & Mathematics with Applications* **63**(6)(2012), 1104–1110.

[4] Toloo, M., Notes on classifying inputs and outputs in data envelopment analysis : a comment, *European Journal of Operation Research* **235**(3)(2014), 810-812.

[5] Scheel, H.: Undesirable outputs in efficiency valuations. *European Journal of Operation Research* **132**(2001), 400–410.

[6] Zhu, J.: *Quantitative models for performance evaluation and benchmarking* : *data envelopment analysis with spreadsheets*. (2nd ed.). Boston, Springer, 2009.

[7] Charnes, A., Cooper, W. W., Seiford, L. and Stutz, J.: Invariant multiplicative efficiency and piecewise Cobb-Douglas envelopments. *Operations Research Letters* **2** (3)(1983), 101–103.

[8] Seiford, L. M., and Zhu, J.: Modeling undesirable factors in efficiency evaluation. *European Journal of Operation Research* **142**(2002), 16–20.

[9] Portela, M. C. A. S., Thanassoulis, E. and Simpson, G.: Negative Data in DEA: A Directional Distance Approach Applied to Bank Branches. *The Journal of the Operational Research Society* **55**(10)(2004), 1111–1121.

[10] Sharp, J. A., Meng, W. and Liu, W.: A modified slacks-based measure model for data envelopment analysis with 'natural' negative outputs and inputs. *The Journal of the Operational Research Society* **58**(12)(2007), 1672–1677.

[11] Emrouznejad, A., Anouze, A. L., and Thanassoulis, E.: A semi-oriented radial measure for measuring the efficiency of decision making units with negative data, using DEA. *European Journal of Operation Research* **200**(1)(2010), 297–304.

[12] Cheng, G., Zervopoulos, P., and Qian, Z.: A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis. *European Journal of Operation Research* **225**(1)(2013), 100–105.

[13] Allahyar, M., and Rostamy-Malkhalifeh, M.: Negative data in data envelopment analysis: Efficiency analysis and estimating returns to scale. *Computer & Industrial Engineering* **82**(2015), 78–81.

[14] Toloo, M. Barat, M. and Masoumzadeh, A.: Selective measures in data envelopment analysis. *Annals of Operations Research* **226**(1)(2015), 623–642.

[15] Toloo, M., and Tichý, T.: Two alternative approaches for selecting performance measures in data envelopment analysis. *Measurement* **65**(2015), 29–40.

[16] Chambers, R. G., Chung, Y., and Färe, R.: Benefit and Distance Functions. *Journal of Economic Theory* **70**(2)(1996), 407–419.

[17] Chambers, R. G., Chung, Y., and Färe, R.: Profit, Directional Distance Functions, and Nerlovian Efficiency. *Journal of Optimization Theory and Applications* **98**(2)(1998), 351–364.

[18] Toloo, M., Selecting and full ranking suppliers with imprecise data: A new DEA method, *International Journal of Advanced Manufacturing Technology* **74**(5-8)(2014), 1141–1148.

[19] Toloo, M., Ertay, T., The most cost efficient automotive vendor with price uncertainty: A new DEA approach, *Measurement* **52**(1)(2014), 135–144.

[20] Toloo, M., Aghayi, N., Rostamy-malkhalifeh, M., Measuring overall profit efficiency with interval data, Applied Mathematics and Computation 201(1–2)(2008), 640–649.

Defense expenditure and economic growth in Visegrad group countries: a panel data analysis

Tereza Ambler[1], Jiří Neubauer[2]

Abstract. The defense economics literature reveals the complexity and inconsistency of examining the relationship between military spending and economic development. The current scientific research has not established a clear approach to handling this large problem. The defense expenditure has been described and investigated by political, strategic and economic factors. This paper aims to analyze the relationship between military expenditure and economic growth in Visegrad group countries (V4) Czech Republic, Slovak Republic, Hungary and Poland. In this study, we focus on trends of military expenditure and some macroeconomic factors which are influencing military expenditure such as the growth rate of GDP, balance of payments, inflation, foreign direct investment, government debt or net lending/borrowing. The effects of determinants of military expenditure are examined with time series data of the period 1995–2015. The empirical model is based on the statistical method of regression and panel regression.

Keywords: military expenditure, economic growth, V4 countries, panel data analysis.

JEL classification: C33, E69
AMS classification: 62J05

1 Introduction

A closer examination of the relationship between military spending and economic growth has allowed the development of economic mathematical methods in the second half of the twentieth century. The era of mathematical and statistical analysis in the field of defense economics has started with Benoits paper [2] investigating the impact of military expenditure on economic growth in 1973. Since then, various approaches have been applied examining this relationship. With the beginning of the Cold War and changes of the power distribution military spending was favored at the expense of non-defense spending [6]. The disintegration of the bipolarity after 1990 has caused the reduction of military budgets and the effort to replace the shortage of funds by more efficient use of resources. Coulomb and Fontanel [3] investigated the possibility of strengthening the military strategic position on France's case despite the reduction in military spending after 1990. Thus, military expenditure has become a variable adjusting to the current economic situation. In the post-cold period, the cuts in military spending were expected and seemed to be inevitable. Odehnal and Neubauer [16] confirmed this tendency in the study of Germany, Great Britain and France. Still, there were some countries excluded from this trend [5], different determinants of military spending were detected even in NATO countries [17]. In this article, we focus our attention on possible determinants of military expenditure in V4 countries of the period 1995–2015 to examine their effects after the end of Cold War.

2 Data and models

2.1 Data

Analysis of determinants of military expenditure in the Czech Republic (CZE), Slovakia (SVK), Poland (POL) and Hungary (HUN) are based on time series of predominantly macroeconomic variables in time period 1995–2015 (the database SIPRI, The World Bank). To describe military expenditure – MILEX [% of a gross domestic product (GDP)], we use following variables: economic growth – EG [%] , foreign direct investment – FDI [% of GDP], inflation – INFL [%], balance of payments – BOP [billions of dollars], population – POP [in millions], net lending/borrowing – NLB [% of GDP], government debt – DEBT [% of GDP].

2.2 Panel data models

The panel data approach belongs to several statistical methods frequently used to describe relationship among macroeconomics variables in last decades, for example, a question of main inflationary factors in V4 countries is

[1]University of Defence, Department of Economics, Kounicova 65, Brno, Czech Republic, intedo@seznam.cz
[2]University of Defence, Department of Econometrics, Kounicova 65, Brno, Czech Republic, jiri.neubauer@unob.cz

discussed in the paper [19], issues connected with military expenditure modeling are examined in the papers [1], [7], [11] or [21].

The general panel model used in econometrics is

$$y_{it} = \alpha_{it} + \beta'_{it} X_{it} + u_{it},$$

where $i = 1, 2, \ldots, n$ is the individual index (group, country), $t = 1, 2, \ldots, T$ is the time index and u_{it} is a random zero mean disturbance term, X_{it} is a $k \times 1$ vector of independent variables, β_{it} is a $k \times 1$ vector of parameters. The parameters β_{it} are not estimable with $N = n \times T$ data points, therefore a number of assumptions are usually made about the parameters, the errors and the exogeneity of regressors. Let us assume that $\alpha_{it} = \alpha$ and $\beta_{it} = \beta$ for all i, t. We get the model

$$y_{it} = \alpha + \beta' X_{it} + u_{it},$$

which is a standard linear model *pooling* all data across i and t, it can be estimated by ordinary least squares (OLS).

To model individual heterogeneity, let us assume that the error term has two separate components $u_{it} = \mu_i + \epsilon_{it}$, where μ_i is specific to the individual and does not change over time.

$$y_{it} = \alpha + \beta' X_{it} + \mu_i + \epsilon_{it}$$

The error term ϵ_{it} is usually assumed independent of both the regressors X_{it} and the individual component μ_i. If the individual component is correlated with the regressors, it it customary to treat the μ_i as next n parameters to be estimated. This is called the *fixed effect* model [9]. If we denote $\alpha_i = \alpha + \mu_i$ we obtain the model

$$y_{it} = \alpha_i + \beta' X_{it} + \epsilon_{it}.$$

This model is sometimes called the *least squares dummy variable* model, it is usually estimated by OLS.

If the individual component μ_i is uncorrelated with the regressors, the model is termed *random effect*, μ_i are not treated as fixed parameters, but as random drawings from a given probability distribution [9]. To get greater efficiency, generalizes least squares (GLS), taking into account the covariance structure of error term, may be used.

3 Empirical results

As a first step, we estimate a linear regression model for each country. The results are summarized in table 1. The table contains estimates for full models (all variables are taken in account), and the reduced models only with statistically significant parameters. It should be noted that economic growth and foreign direct investment are not significant in any country. Inflation is significant only for the Czech Republic (with a negative sign), balance of payments is significant (with a positive sign) in all countries except for Hungary. There is a significant negative effect of population in the Czech Republic and Slovakia, whereas the effect is positive in Poland and Hungary. Net lending/borrowing plays a significant negative role in the Czech Republic, Slovakia and Poland like government debt in all countries except for Poland. The question of spurious regression is closely connected with linear regression models applied to time series when some of them are non-stationary. Unit root tests, such as augmented Dickey-Fuller (ADF) or Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, are usually used to detect non-stationary behavior of analyzed time series. In case that some of them are non-stationary, it is necessary to test whether examined time series are cointegrated. It should be emphasized that times series are rather short (21 observations). The power of unit root tests and tests of cointegration for such a short time series is small. Time series of military expenditure, population and payments balance can be considered non-stationary, variables net lending/borrowing and government dept seem to be stationary (except for government debt in Hungary). We use the significance level 0.05 in all testing procedures. Normality of residuals of estimated model was not rejected (Shapiro-Wilk test, Lilliefors test), residuals were found uncorrelated (Ljung-Box test). There are several approaches to cointegration testing. One is based on a single equation method (Engle-Granger test), another employs a vector error correction model (Johansen tests). According the results of Johansen tests, the systems of analyzed time series in each country are cointegrated. However, Engle-Granger tests do not reject the null hypothesis of no cointegration. This test is based on ADF test applied on residuals from linear regression models. The power of the test for short time series (and this is our case) is very small. As a result, the tests do not reject the existence of a unit root in residuals. On the other hand, the autocorrelation functions of residuals do not contain significant values, especially for lag 1, when for non-stationary (integrated) time series the value close to one is expected. We came to conclusion that the regressions are not spurious.

The aim of this contribution is to describe military expenditure of V4 countries by selected variables using panel data models. We estimated pooling, fixed and random effect model. Estimation results are summarized in table 2. Statistical software R (the package *plm* [4]) and Gretl was used for necessary calculation. The table contains, as with the individual regression models, estimates for full and reduced (final) models. Economic growth was found

	Dependent variable: MILEX							
	CZE (full)	CZE	SVK (full)	SVK	POL (full)	POL	HUN (full)	HUN
EG	−0.008		0.029		0.044		−0.012	
	(0.007)		(0.020)		(0.042)		(0.013)	
FDI	0.007		−0.035		−0.045		−0.003	
	(0.008)		(0.025)		(0.045)		(0.003)	
INFL	−0.019*	−0.020**	0.026		0.002		0.006	
	(0.009)	(0.009)	(0.024)		(0.010)		(0.012)	
BOP	0.014*	0.015*	0.108***	0.148***	0.010**	0.009**	−0.016	
	(0.008)	(0.008)	(0.035)	(0.026)	(0.004)	(0.003)	(0.014)	
POP	−1.692***	−1.672***	−15.412**	−15.957***	0.560*	0.988***	0.248	0.746***
	(0.254)	(0.226)	(7.052)	(5.177)	(0.305)	(0.135)	(0.678)	(0.178)
NLB	−0.023***	−0.025***	−0.060**	−0.051**	−0.057**	−0.044**	−0.012	
	(0.007)	(0.007)	(0.021)	(0.019)	(0.023)	(0.020)	(0.016)	
DEBT	−0.017***	−0.017***	−0.029**	−0.040***	−0.016		−0.015***	−0.014***
	(0.004)	(0.003)	(0.012)	(0.009)	(0.010)		(0.005)	(0.003)
Const.	19.497***	19.307***	85.703**	89.298***	−19.083	−36.089***	−0.329	−5.294**
	(2.613)	(2.270)	(37.740)	(27.652)	(11.987)	(5.169)	(6.930)	(1.883)
Observations	21	21	21	21	21	21	21	21
R^2	0.972	0.967	0.905	0.870	0.886	0.861	0.875	0.832
Adjusted R^2	0.957	0.956	0.853	0.838	0.825	0.837	0.808	0.813
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 1 Linear regression models for individual countries, standard errors are in parenthesis; MILEX – military expenditure, EG – economic growth, FDI – foreign direct investment, INFL – inflation, BOP – balance of payments, POP – population, NLB – net lending/borrowing, DEBT – government debt

significant only for the pooling model. In this model, almost all variables are statistically significant, only variable FDI is not significant. The estimated parameters in the final fixed and random models are nearly the same. We conducted several statistical tests to decide which model is better for the description of military expenditure. The first – the test of poolability – tests the hypothesis that the same coefficients apply to each individual (country). The p-value of this test is 0.00029 which means, that the fixed or random effect model should be preferred to the pooling model. Hausman test [8] offers the way how to compare fixed and random effect models. The p-value of this test for final models is 0.49251. According to this result we favor the random effect model

$$\mathit{MILEX} = 2.458 + 0.017 \cdot \mathit{INFL} + 0.011 \cdot \mathit{BOP} - 0.037 \cdot \mathit{NLB} - 0.024 \cdot \mathit{DEBT}.$$

From the perspective of the V4 countries, military expenditure is mainly influenced by inflation, balance of payments, net lending/borrowing and government debt. The effect of inflation and balance of payments is positive meaning that of inflation or balance of payments is increasing, military expenditure is increasing as well and vice versa. The opposite effect can be detected for the variables net lending/borrowing and government debt.

Analysis of residuals reveals that estimated model is not able to describe behavior of military expenditure in Slovakia for the years 1995 and 1996 satisfactorily. These values are considerably higher than other values, and corresponding residuals are big which causes rejection of normality. If one trims these values, normality of residuals is not rejected by common normality tests (Shapiro-Wilk, Lilliefors, Jarque-Bera test). The question of spurious regression is relevant in context of panel data modeling too. Firstly, we conducted unit root tests for panel data (Levin-Lin-Chu [13] and Im, Pesaran and Shin test [10]) with the result that time series of military expenditure, payments balance a government debt are non-stationary, inflation and net lending/borrowing are stationary. Cointegration relationship can be tested by the panel cointegration tests proposed by Pedronni [18], package *pco* in R software [15]. All tests reject the null hypothesis of no cointegration, which means that the system of analyzed panel data is cointegrated.

	Dependent variable: MILEX					
	pooling (full)	pooling	FE (full)	FE	RE (full)	RE
EG	0.034***	0.033***	0.014		0.014	
	(0.012)	(0.011)	(0.011)		(0.011)	
FDI	0.002		0.002		0.001	
	(0.004)		(0.004)		(0.004)	
INFL	0.017***	0.018***	0.019***	0.016***	0.016***	0.017***
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)
BOP	0.012**	0.011**	0.014**	0.012**	0.013**	0.011**
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
POP	0.012***	0.012***	−0.234		−0.047	
	(0.003)	(0.003)	(0.229)		(0.111)	
NLB	−0.055***	−0.055***	−0.039***	−0.037***	−0.040***	−0.037***
	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
DEBT	−0.014***	−0.014***	−0.024***	−0.026***	−0.024***	−0.024***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Const.	1.597***	1.602***			3.152	2.458***
	(0.123)	(0.122)			(2.437)	(0.221)
Observations	84	84	84	84	84	84
R^2	0.620	0.619	0.639	0.626	0.635	0.612
Adjusted R^2	0.585	0.590	0.590	0.592	0.601	0.592
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 2 Panel data models – pooling, fixed effect (FE) and random effect (RE) models, standard errors are in parenthesis; MILEX – military expenditure, EG – economic growth, FDI – foreign direct investment, INFL – inflation, BOP – balance of payments, POP – population, NLB – net lending/borrowing, DEBT – government debt

4 Discussion

The aim of this study has been to examine the relationship between military expenditure and economic growth in Visegrad group countries (V4) Czech Republic, Slovak Republic, Hungary and Poland. In this complex problem we decided to focus mainly on the following variables which could determine military expenditure such as the growth rate of GDP, population, balance of payments, inflation, foreign direct investment, government debt or net lending/borrowing. The results of the estimated panel model show that balance of payments and inflation have a positive impact on military expenditure, and that net lending/borrowing and government debt have a negative impact on military expenditure. For the whole group, we find no impact of economic growth and FDI on military expenditure.

The growth rate of GDP was chosen as an indicator of the economic growth. According to Tambudzai [20] increasing levels of GDP causes rising of the demand of defence to protect natural and national resources. Kollias et al. [12] used the FDI inflows to Cyprus as approximation of the safeness of national economy. Even though the foreign direct investment was important for this study of Cyprus, in our results we did not observe any significant influence across the whole sample. This may indicate that the use of the FDI variable is appropriate for developing economies, see [12], [20]. The impact of military expenditures on external debt and external borrowing was investigated by Dunne, Perlo-Freeman and Soydan [7]. The results indicate that military expenditures have positive impact on debt for developing countries. For group of V4 we discovered reversed effects of the government debt or net lending/borrowing on military expenditures. Net lending/borrowing had a significant negative effect on military expenditures in the Czech Republic, Slovakia and Poland. Government debt had a significant negative effect in the Czech Republic, Slovakia and Hungary.

The change of population can possibly affect military expenditure as a part of government expenditure. Tambudzai and Harris [20] saw ambiguous effects in the influence of population on military expenditure. Increasing population is causing an added need of protection against an external threat. On the other hand, rising size of population might have a crowding out effect on military spending because of demanding social and health services.

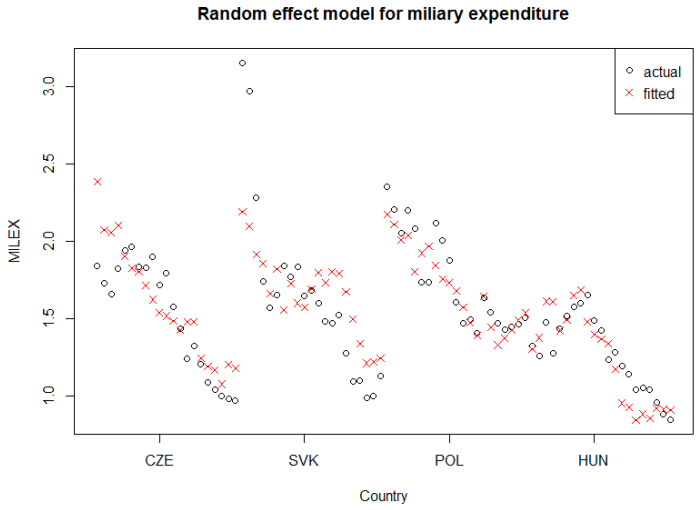


Figure 1 Random effect model for military expenditure

The negative effect of population of the Czech Republic and Slovakia on military expenditures suggests that social government expenditures are declining the military expenditures. An exactly opposite effect of population was discovered in Hungary and Poland. The inflation is negatively significant only for the Czech Republic according to our findings. For the rest of the countries we did not find any relevant effects.

The balance of payments was used here as an indicator of the economy's transactions with the rest of the world. Dunne and Nikolaidou [5] used the trade balance as one of the economic variables in their study of Greece. The trade balance as a part of the balance of payments can be also used as an external factor of military expenditures expressing the openness of the national economy and showing the trend of the balance of payments. Due to no evidence of any impact of economic variables on military expenditure in Greece, Dunne and Nikolaidou did not find any impact of trade balance on military expenditure. The balance of payments was used here as an indicator of the economys transactions with the rest of the world. Dunne and Nikolaidou [5] used the trade balance as one of the economic variables in their study of Greece. The trade balance as a part of the balance of payments can be also used as an external factor of military expenditures expressing the openness of the national economy and showing the trend of the balance of payments. Due to no evidence of any impact of economic variables on military expenditure in Greece, Dunne and Nikolaidou did not find any impact of trade balance on military expenditure. Our finding is that the balance of payments is having a significant and positive impact on military expenditure on the Czech Republic, Slovak and Poland.

5 Conclusion

The contribution deals with panel data models for V4 countries with the aim to describe behavior and possible determinants of military expenditure in these countries. Several mainly macroeconomic indicators were employed and the model was estimated. The random effect model with four regressors (inflation, balance of payments, net borrowing/lending and government debt) was finally used. First, the empirical results do not prove, for the observed four countries as a whole, any impact on economic growth, FDI and the size of population on military expenditure. Second, the panel data model indicates that, an increase of the balance of payments and inflation has positive effects on military expenditures. Third, there is a negative impact of government debt and net lending/borrowing on military expenditures for V4 countries. The balance of payments was found to have a positive effect on military expenditures for all examined countries expect Hungary. Military expenditures of each V4 country are positively influencing the balance of payments. The empirical results suggested that military expenditures are mainly influenced also by inflation, government debt and net lending/borrowing for V4 countries. We have identified various effects of other indicators on military expenditures. Inflation is significantly negative only for the Czech Republic (no significance shown for the rest of the countries). There is a significant negative effect of the population in the Czech Republic and Slovakia, while the effect is positive in Poland and Hungary. Government debt is significantly negative for the Czech Republic, Slovakia and Hungary. Net lending/borrowing is significantly negative for the Czech Republic, Slovakia and Poland.

Acknowledgements

The contribution is supported by the grant No. SV17-FVL-K102-AMB of the Ministry of Education, Youth and Sports, Czech Republic.

References

[1] Ali, H. E.: Military Expenditure and Inequality in the Middle East and North Africa: a Panel Analysis, *Defence and Peace Economics* **23**, 6 (2012), 575–589.

[2] Benoit, E.: *Defence and Economic Growth in Developing Countries*. Lexington Books, Boston, 1973.

[3] Coulomb, F. and Fontanel, J.: An Economics Interpretation of French Military Expenditures, *Defence and Peace Economics* **16**, 4 (2005), 297–315.

[4] Croissant, Y. and Millo, G.: Panel Data Econometrics in R: The plm Package, *Journal of Statistical Software* **27**, 2 (2008). URL http://www.jstatsoft.org/v27/i02/.

[5] Dunne, P. and Nikolaidou, E.: Military expenditure and economic growth: A demand and supply model for Greece, 1960–96, *Defence and Peace Economics* **12**, 1 (2001), 47–67.

[6] Fontanela, J. and Hébert, J. P.: The end of the French grandeur policy, *Defence and Peace Economics* **8**, 1 (2007) , 37–55.

[7] Dunne, J. P., Perlo-Freeman, S. and Soydan, A.: Military expenditure and debt in small industrialised economies: A panel analysis, *Defence and Peace Economics* **15**, 2 (2010), 125–132.

[8] Hausman, J. A.: Specification Tests in Econometrics, *Econometrica* **46**, 6 (1978), 1251–1271.

[9] Hsiao, Ch. *Analysis of Panel Data*. 3rd edition. Cambridge University Press, 2014.

[10] Im K. S., Pesaran, M. H. and Shin Y.: Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics* **115**, 1 (2003) , 53–74.

[11] Kollias, Ch., Mylonidis, N. and Paleogou, S. M.: A Panel Data Analysis of the Nexus between Defence Spending and Growth in the European Union, *Defence and Peace Economics* **18**, 1 (2007), 75–85.

[12] Kollias, Ch., Naxakisb, Ch. and Zarangasb, L.: Defence Spending and Growth in Cyprus: A Causal Analysis, *Defence and Peace Economics* **15**, 3 (2004), 299–307.

[13] Levin, A., Lin, C. F. and Chu, C. S. J.: Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties, *Journal of Econometrics*, **108**, 1 (2002), 1–24.

[14] Looney, R. E. and Frederiksen, P. C.: The economic determinants of military expenditure in selected East Asian Countries, *Contemporary Southeast Asia* **11**, 4 (1990), 265–277.

[15] Marinov, G.: *pco: Panel Cointegration Tests*. R package version 1.0.1., 2015. https://CRAN.R-project.org/package=pco

[16] Odehnal, J. and Neubauer, J.: Economic Determinants of Military Spending – Causal Analysis, *Ekonomický časopis* **63**, 10 (2015), 1019–1032.

[17] Odehnal, J. Sedlačík, M.: The Demand for Military Spending in NATO Member Countries. In: *Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2014 (ICNAAM-2014), AIP Conference Proceedings, 1648.* American Institute of Physics, Melville, New York, 2015.

[18] Pedroni, P.: Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors, *Oxford Bulletin of Economics and Statistics* **61**, S1 (1999), 653–670.

[19] Roubalová, L. and Hampel, D.: The main inflationary factors in the Visegrad Four. In *34th International Conference Mathematical Methods in Economics (MME 2016*). (2016), 741–746.

[20] Tambudzai, Z. and Harris, G.: *What drives military expenditure? Evidence from Zimbabwe and the SADC region.* Saarbrucken: LAP Lambert Academic Pub, 2011.

[21] Yildirim, J, Sezgin, S. and Öcal, N.: Military Expenditure and Economic Growth in Middle Eastern Countries: a Dynamic Panel Data Analysis, *Defence and Peace Economics* **16**, 4 (2005), 283–295.

Does Relative Income Have an Impact on the Consumption of Hungarian Households?

Ondřej Badura[1]

Abstract. This paper examines the influence of relative income on the consumption function, particularly the issue of the precise quantification of the relationship between the marginal propensity to consume and the position of the household in the income distribution represented just by relative income. Mainstream theory of consumption, though it is aware of this effect, for its focus primarily on the aggregate consumption, it pays virtually no attention to this. However, if it is possible to accurately describe and quantify this so-called relative income effect, it will mean a relevant contribution to the explanation of individual consumption patterns or consumption of income categories. The aim of this work is to find and precisely quantify the impact of relative income on the marginal propensity to consume on the example of Hungarian households and thus to support the idea of interdependent concept of utility and consumption. To achieve this goal, we use a panel regression analysis. The results unambiguously confirm the statistical validity of the initial hypothesis, which shows that the consumption decisions of Hungarian households are interdependent. Moreover the more precise form of the wanted regression equation was found in estimation by the polynomial function.

Keywords: Relative income, marginal propensity to consume, interdependent consumption, panel regression.

JEL Classification: D11, D12
AMS Classification: 62P20

1 Introduction

Household consumption, which is usually the motivation and the final goal of any economic effort, it is then logically also the largest component of GDP and hence the most significant driving force of economic growth. It is not surprising that the development of theory and the empirical research of consumption is then concentrated primarily on the aggregate scale. This fact eventually resulted in a situation corresponding to the current state of knowledge, and thus that many properties and principles of aggregate consumption function are relatively well theoretically described and empirically tested and proved. On the other hand, of course, there are also characteristics that are non-testable or insignificant at the aggregate level and their influence is growing more strongly, the closer we are to the very nature of the economic microcosm. These phenomena have often been underestimated, hidden in the shadow of macroeconomic research. One of these effects is also the decreasing value of the marginal propensity to consume across the distribution of disposable income.

The approach of lifecycle and permanent income theory, which has become a baseline model for further economic research, it is based on strong microeconomic fundamentals, but as such it leads primarily to the consumption function of aggregate and therefore hypotheses that can be tested at a macro level. Although the variability of the marginal propensity to consume across different income types of households is intuitive and self-evident also from the perspective of LC-PIH (life cycle - permanent income hypothesis)[2], for its macroeconomic focus, this mainstream approach does not virtually deal with it. Also later concepts such as myophia and liquidity constraints by Flavin [6] or buffer stock saving model by Carroll [4] doesn't change anything in that direction. The theoretical explanation of this phenomenon, which basics can be found in the work of Velben [13], can be sought out of the mainstream theory, as shown, for example, by Ackerman [1]. The hypothesis of relative income by Duessenberry [5] nowadays newly reformulated for example by Palley [12] or Alvarez-Cuadrado and Long [2] seems to be the most ambitious one. It represents a consumer concept based on an interdependent concept of utility, which principles and conclusions, if they become verified, can be a welcomed contribution and

[1] VŠB - Technical University of Ostrava, Department of Economics, Sokolská třída 33, 702 00 Ostrava, Czech Republic, e-mail: ondrej.badura@vsb.cz

[2] The theoretical approach to consumption based on the original works: Modigliani and Brumberg [11] and Friedman [7]. In addition to the element of rational expectation, we can refer to the so-called random walk model as defined by Hall [8].

enrichment of the current state of knowledge about the patterns of individual consumption or consumption of income categories.

The aim of this work is to find and precisely quantify the impact of relative income on the marginal propensity to consume on the example of Hungarian households using panel regression and thus to support the idea of interdependent concept of utility and consumption.

2 Methods and data

2.1 Methods

It is worth pointing out at this point that the main motive of this work is to prove the influence of the relative income of households on the value of their marginal propensity to consume by formulating the particular form of a possible functional relationship. Therefore the concept of relative income remains the key matter for us. From the principle point of view, it is de facto an quantification and therefore the possibility of mathematical and economic formulation of the issue of the position of the household within the distribution of disposable income. From a definition point of view, it is the ratio of disposable income (YD) to its weighted average throughout society, as shown in Equation 1:

$$YRD = \frac{YD}{\overline{YD}} \tag{1}$$

Now, just the variable in the denominator has to be specified. The theoretical background, which is given us by a reference of James Duesenberry's work, leaves us relative freedom in this area, so we need to consider the appropriateness and the most suitable explanatory ability of the chosen definition. For the purposes of further analysis, the variable $\overline{YD}$ was defined as the wide-weighted average of YD throughout population, where the weights were set by the average number of household members in a given income category:

$$\overline{YD} = \frac{\sum_{i=1}^{n} YD_i \cdot w_i}{\sum_{i=1}^{n} w_i} \tag{2}$$

where YD_i characterizes the income of the i-th income category, w_i the average number of household members in the i-th income category and *n* the number of those categories. The determination of weights was the central issue here and our definition best corresponds to the original concept and meaning of the relative income effect.

For the attempt of expressing the particular form of predicted functional dependence we use a panel regression analysis, mainly because of the limited number of statistically measured income categories (ie a small number of observations). The form of the general equation of the wanted one-dimensional linear regression model will depend on whether the method of fixed or random effects is used in panel regression. Which of those methods is more appropriate for our data, it will be shown by Hausman's test in later part of analysis, so it is now necessary to consider both variants. In the case of using of fixed effects, the regression equation would be given by:

$$MPC_{i,t} = \alpha_i + \beta \cdot YRD_{i,t} + u_{i,t} \tag{3}$$

where $MPC_{i,t}$ - the marginal propensity to consume for the i-th category at time *t* is expressed by the α_i -level constant for the i-th income category, the product of $YRD_{i,t}$ - the relative disposable income for the i-th category at the time *t* and of the regression coefficient β expressing the sensitivity of the marginal propensity to consume on the relative disposable income. The variable $u_{i,t}$, then represents the random component.

In the case of a regression estimation based on random effects, the wanted relationship would be characterized more simply and unambiguously in the form of:

$$MPC_{i,t} = \alpha + \beta \cdot YRD_{i,t} + u_{i,t} + \varepsilon_{i,t} \tag{4}$$

where newly α represents a level constant for all categories, $u_{i,t}$ random component between the categories and $\varepsilon_{i,t}$ random component within the income category.

Either way, a negative value of the β coefficient is an important premise for all possible resulting panel regression variants, because according to the principles of the Duesenberry's hypothesis with the increasing relative disposable income, the marginal propensity to consume must necessarily decline.

2.2 Data

The assumption of negative linear dependence of MPC on YRD will be tested here on the example of aggregate data for the household budget situation in Hungary, that's why all input data for this analysis were taken from the database of Hungarian Central Statistical Office [9]. The original input data are given by the annual statistics between 2010-2015, which are basically set into two time series, which are further divided into ten other subfolders. The followed 6 observations are therefore basically written in two variables:

YD - average annual nominal household disposable income per capita in HUF,

C - average annual nominal household consumption per capita in HUF.

As we can see, we work here with average data per capita. For a better demonstration of the validity of the Duesenberry's hypothesis, this procedure is certainly more appropriate. An important matter is also the already mentioned secondary subdivision of basic variables. The YD and C indicators are equally divided into ten other subfolders reflecting the income and consumption situation of each household types sorted by deciles ascendingly according to disposable income. We record here, as a result, 20 input time series, divided into 10 panels by the types of income categories. The indicators directly entering the subsequent panel regression are YRD, calculated according to formula 1 and 2 and APC expressed by:

$$APC = \frac{C}{YD} \tag{5}$$

It is necessary to realize at this point that we work here with income categories (not with individual households) for which the APC value is independent of YD and in the absence of an intercept it is at every point equal to MPC. That's why we could use this simple equivalence here, where values of MPC are substituted by the average propensity to consume. Finally we note that the original input data in this study are given by a nominal expression of consumption and disposable income, but due to the relative nature of the MPC and YRD indicators, the undesirable effect of changes in price levels is to be completely canceled out anyway.

3 Results

First, the stationarity of the data directly entering the panel regression was tested. According to Levin-Lin-Chu unit root test, the panels MPC and YRD are stationary, at least at 5% significance level, which we work with. The same result in stationarity verification was also achieved by the Harris-Tzavalis test.

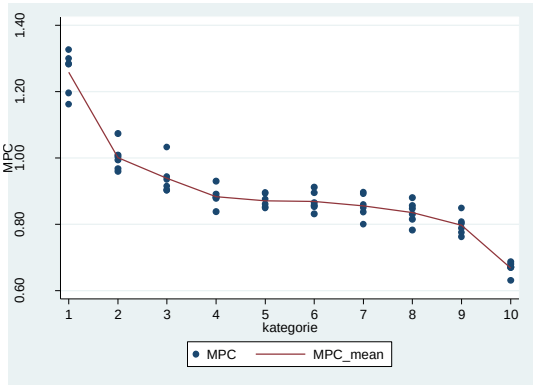


Figure 1 Development of MPC across income categories

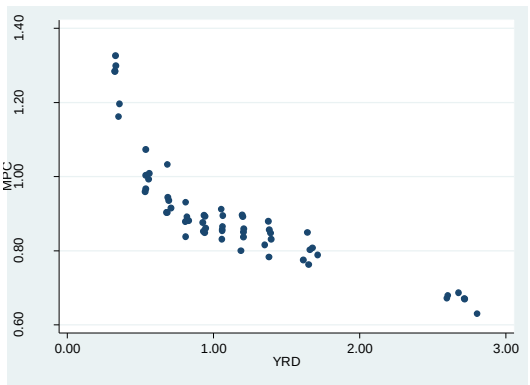


Figure 2 Visual assessment of linear dependency

Figure 1 illustrates the development of the marginal propensity to consume across income categories. At first glance, it is evident that the value of MPC isn't constant in any of the categories and so it deflects in time. This implies, according to the principles of the Duesenberry's hypothesis, that the magnitude of the force of the relative income effect is not constant over time. However, a more significant fact, resulting from Figure 1, is that the average value of MPC actually falls across income categories, as predicted by the relative income hypothesis.

Figure 2 then serves as a preliminary visual assessment of predicted dependence. There is evident, at first sight that the relationship of MPC and YRD is not entirely linear in its foundations, but it is rather convex in its characteristics. Figure 2 thus predicts that linear function would probably not be a suitable approximation tool in this case, and alternative estimates using higher degree polynomial functions could be a significant benefit for the explanatory ability of the model.

Before it can be proceeded to the final estimation of the regression parameters of the wanted dependence, it is necessary to determine whether it should be used the method of fixed or random effects, in other words,

whether there are differences between the categories so significant, that they should be captured in a separate level constant for each category. This dilemma is unambiguously judged by the Hausman test and its very low value of calculated statistics H = 0.03, which indicates that a suitable panel regression here is the method of random effects at virtually any level of significance. In this case, it is also worthwhile checking whether there is at least a statistically significant variance between the categories, in other words whether it is necessary to use panel regression or just simple OLS. Solving of this problem is offered by Breusch-Pagan LM test with a value of test statistic 99.7, which means that our data show a panel effect – ie a non-zero variance between the entities.

Number of observation	60
χ^2	12.68
P-value	0.0004
$\mathbf{R^2}$ **within categories**	0.0295
$\mathbf{R^2}$ **between categories**	0.7054
$\mathbf{R^2}$ **overall**	0.67

Table 1 Final estimation of linear function according to equation 4, part 1

MPC	**Coefficient**	**Robust standard error**	**z**	**P-value**
YRD	- 0.1937	0.0544	- 3.56	0.000
α	1.1169	0.0808	13.82	0.000

Table 2 Final estimation of linear function according to equation 4, part 2

The results of the final panel regression using random effects are summarized in Tables 1 and 2. In this final estimation of the desired functional pattern, a robust method of estimating the standard error through the White estimator was used, thus the model was treated against possible autocorrelation and heteroskedasticity. There is an important finding of relatively high value of R^2, which indicates that 67% of the variability of MPC was explained by YRD, where this result was mainly achieved by the effect of the relative income between the categories. This is entirely in line with the initial premise, since "keeping up with the Joneses" effect should be seen primarily in comparing different income categories, not so much within one category. Either way, a relatively high value of R^2 (relatively to the use of only one explanatory variable) thus only confirms the validity of the initial assumption about the influence of relative disposable income on the marginal propensity to consume. The model itself is no doubt statistically significant as well as its regression coefficient. The wanted regression coefficient β achieves a negative value (as we expected) that can not be affected nor by its standard deviation. The resulting model therefore corresponds to the underlying economic theory and predicts that for the income categories of Hungarian households it applies, that the change in the relative disposable income by 0.1 also changes the value of the marginal propensity to consume by about 0.02 in the opposite direction.

We have found a possible particular form of the relationship between MPC and YRD. Now we have just left to verify that the linear function is indeed a suitable approximation of the observed dependence. Figure 2 already indicates that the dependence between MPC and YRD doesn't seem to be much linear. But let's not just rely on the visual assessment and let's try to estimate the relationship by more complex equation. For this purpose we use higher polynomial functions from 2 to 7 degrees, which allow for diverse variations of the regression curve and at the same time they remain linear in their estimated parameters.

Polynomial degree	$\mathbf{R^2}$ **overall**	χ^2 **statistics**	**P-value**	**Number of significant β coefficients**
1	0.67	20.96	0.0000	1
2	0.8034	41.01	0.0000	2
3	0.9135	157.79	0.0000	3
4	0.9431	911.05	0.0000	4
5	0.9549	1143.4	0.0000	5
6	0.9552	1129.15	0.0000	2
7	0.9549	-	-	3

Table 3 Estimates of polynomial functions (basic, non-robust estimates)

Table 3 summarizes the most important characteristics of the models estimated by the polynomial functions from the first to the seventh degree, where the first degree polynomial is de facto the originally used simple linear function and serves for comparison of obtained results. In all cases, the random effect method was used. However, it should be emphasized at this point that the individual regression parameters of the polynomial higher than the first degree lose their economic interpretation, so Table 3 does not deal with them and their signs and magnitudes will not be discussed anymore. Apart from the fact that all estimated models are no doubt statistically significant, we are most interested in the value of the coefficient of determination, respectively in its change. As we can see, this increase is quite considerable. Highest values of R^2 is reached by the polynomial of the 6th degree, which also means an increase of the explained variability of MPC by 28.52 percentage points compared to the initial model. While the complexity of the approximation function has increased considerably, the very high value of the resulting R^2 should be sufficient compensation in this case. Economically speaking, the yields stemming from the significantly higher explanatory ability of the model are probably higher than the costs associated with breaking the simplicity and elegance of a simple linear function. The number of statistically significant coefficients in estimation can also be used as another tool for decision-making. We can see that, until the polynomial of the 5th degree, their number grow equivalently to the polynomial degree, but subsequently their number decreased rapidly. This means that, the explained variability of the polynomials of 6th and 7th degree, is achieved also through some coefficients, which will probably not be part of the resulting regression equation, so the actual R^2 of that models can be expected to be slightly smaller. And since the increase of the coefficient of determination in the 6th degree polynomial is very small compared to the previous polynomial estimate and in the case of 7th polynomial the change of R^2 is even negative, it is possible to refer to the 5th degree polynomial with the full number of statistically significant coefficients as the most appropriate approximation equation as shown in Table 4 for the final robust estimation.

MPC	Coefficient	Robust standard error	z	P-value
YRD	- 4.9114	0.3352	- 14.65	0.000
YRD^2	6.5087	0.5635	11.55	0.000
YRD^3	- 4.1588	0.4131	- 10.07	0.000
YRD^4	1.2565	0.1367	9.19	0.000
YRD^5	- 0.1446	0.0167	- 8.64	0.000
α	2.3184	0.0642	36.12	0.000

Table 4 Final estimation of the 5th degree polynomial

Finally, we emphasize that result of Hausman's test has significantly influenced the prediction ability of the resulting model. The resulting use of the random effect method means that the regression relationship between MPC and YRD can be expressed generally and elegantly in one single equation, and it doesn't depend on the type of income category or the time we consider. This fact on the other hand has been redeemed by a more complicated interpretation of the coefficient β, which now includes both an inter-category effect and the effects viewed within a single entity. However, as already mentioned above, the interpretation of regression coefficients has been complicated much more by the resulting use of the polynomial function, therefore the above-mentioned problem is de facto no longer a problem any more.

4 Conclusion

The primary motive of this work was to find and prove the influence of relative (disposable) income on the value of marginal propensity to consume. In order to achieve this objective, the panel regression was used on the data on the budgetary situation of income categories in Hungary. However, the problem of the actual definition of relative disposable income had to be resolved first. As shown by the previous paragraphs, this goal has been actually met. Now it is no longer a matter of dispute that changes in marginal propensity to consume can be explained by relative income, which also confirms the validity of "keeping up with the Joneses" effect, at least for the case of Hungarian households. And as it is indicated by relatively high value of the R^2 and the statistically significant regression coefficient of the relative income (when approximated by the linear function), this finding could significantly contribute to our knowledge of consumption behavior of different income groups.

Although the simple linear function estimation is sufficient proof of the effect of relative income on the marginal propensity to consume, one of the above-mentioned goals was to approximate this relationship as accurately as possible. Eventually the fifth degree polynomial regression was found to be the most appropriate description of observed causality, which means that the individual regression coefficients de facto lost their economic interpretation, but the predictive ability of the model as a whole thus became significantly stronger. In other

words, although this polynomial relationship is less economically graspable, it much better and accurately describes the real state of economic reality. In this context, it is advisable to compare the results achieved with the study of Badura [3] which, on data from the Czech Republic, concludes that the relationship of MPC and YRD is on one hand dependent on the considered category and even on the exact year, but also its most acquire form would be still purely linear. Therefore, it is clear that the particular form of the relationship between the marginal propensity to consume and relative income is not a matter of generally validity and it probably depends on the national specifics of the given economy. And because the exact form of this relationship is a matter of crucial importance for the effectiveness of economic policy, the quest for its quantification in other various economic environments remains an important motive for extending of this work.

In conclusion it is worthwhile to note that influence of relative income hypothesis is actually much higher that it is commonly understood, as shown for example by Mason [10]. It has many implications and can help to describe the demand formation more appropriately in many ways. But the most important contribution of Dueseberry's work is deeper understanding of income effect itself. So in the final lines we want to emphasize that the mainstream microeconomics distinguishes only between income and substitution effect. Duesenberry's theory, as well as the conclusions of this study require to add further subdivision and so to distinguish between income effect of absolute (directly influencing the level consumption) and relative (indirectly influencing the level consumption through MPC).

Acknowledgements

This paper has been funded by a grant from Students Grant Project EkF, VŠB-TU Ostrava within the project SP2017/110.

References

[1] Ackerman, F.: Consumed in Theory: Alternative Perspectives on the Economics of Consumption. *Journal of Economics Issues* **31** (1997), 651-664.
[2] Alvarez-Cuadrado, F., and Long, N. V.: The Relative Income Hypothesis. *Journal of Economic Dynamics and Control* **35** (2011), 1489-1501.
[3] Badura, O.: An Influence of Relative Income on the Marginal Propensity to Consume – Evidence from Czech Republic. In: *Proceedings of 14th International Scientific Conference: Economic Policy in the European Union Member Countries* (Majerová, I., Kotlánová, E., eds.). Department of Economics and Public Administration of the Silesian University in Opava, Petrovice u Karviné, 2016, 21-33.
[4] Carroll, Ch. D.: Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis. *Quarterly Journal of Economics* **107** (1997), 1-56.
[5] Duesenberry, J. S.: *Income, Saving, and the Theory of Consumer Behavior*. Harvard University Press, Cambridge, 1949.
[6] Flavin, M. A.: Excess Sensitivity of Consumption to Current Income: Liquidity Constraints or Myopia? *Canadian Journal of Economics* **18** (1985), 117-136.
[7] Friedman, M.: *A Theory of the Consumption Function*. Princeton University Press, New York, 1957.
[8] Hall, R. E.: Stochastic Implication of the Life Cycle – Permanent Income Hypothesis: Theory and Evidence. *Journal of Political Economy* **86** (1978), 971-987.
[9] Hungarian Central Statistical Office. *Time series of annual data - Income and living conditions*. HCSO, Budapest, 2016.
[10] Mason, R.: The Social Significance of Consumption: James Duesenberry's Contribution to Consumer Theory. *Journal of Economic Issues* **34** (2000), 553-572.
[11] Modigliani, F. and Brumberg, R.: Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data. In: *Post-Keynesian Economics* (Kurihara, K., eds.). Rutgers University Press, New Brunswick, 1954, 338-436.
[12] Palley, T. I.: The Relative Income Theory of Consumption: A Synthetic Keynes - Duesenberry - Friedman Model. *Review of Political Economy* **22** (2010), 41-56.
[13] Veblen, T.: *The Theory of the Leisure Class: An Economic Study of Institutions*. Macmillian, New York, 1899.

TOPSIS with Generalized Distance Measure GDM in Assessing Poverty and Social Exclusion at Regional Level in Visegrad Countries

Adam P. Balcerzak[1], Michał Bernard Pietrzak[2]

Abstract. Decreasing poverty and social exclusion is treated as one of the most important aims of Europe 2020 plan, which is supported by EU funds and national financing by EU members. Thus, it should be constantly monitored with application of quantitative methods. The problem of poverty is commonly treated as a multivariate phenomenon, characterized with significant regional diversity. As a result, in the research TOPSIS method was used. The analysis was done for Poland, Czech Republic, Slovakia and Hungary at regional level (NUTS 1) with application of Eurostat data for the years 2011 and 2015. In TOPSIS method a choice of metric used for calculation the distance of objects from positive and negative ideas solutions is a key problem. Generalized distance measure GDM is a metric, which can be applied for variables measured on the ratio scale, interval scale, the ordinal scale or the nominal scale. Therefore, it makes the metric universal for economic research based on variables measured in different scales. From methodological perspective, the article presents applicability of the measure GDM in TOPSIS method. The conducted research confirms significant diversity between the analyzed regions.

Keywords: poverty, social exclusion, multiple criteria decision analysis (MCDA), TOSPIS, general distance measure GDM.

JEL Classification: C38, I32
AMS Classification:90B50, 90C29

1 Introduction

The problem of poverty and social exclusion has been in the spotlight of the European authorities for last decades. After the global financial crisis limiting the range of poverty has been declared as one of the aims of Europe 2020 plan [8], which is supported by the European structural funds and national financing by the EU members. Obtaining this objective is crucial for building conditions for sustainable growth and quicker socio-economic convergence with developed economies [27, 17, 33, 34, 22, 16], and improving macroeconomic stability [10]. Thus, the aim of the research is to quantify the phenomenon of poverty and social exclusion for Visegrad countries at regional level(NUTS 1) in the years 2011-2015.

Most of subjects of economic research should be considered as multidimensional phenomena characterized with many aspects [13, 21]. For each aspect a set of diagnostic variables can be selected, which characterize the aspect and enable its proper description and quantification. In that case, a taxonomic measure of development (TMD) can be used, as the TMD can enable to take into account the influence of all determinants of the economic phenomenon under investigation and allows for a synthetic assessment of its level [18, 12; 30]. TMR as a synthetic variable is a commonly used tool in economic research, allowing to describe and analyze the variability of complex multidimensional economic phenomena and assess their changes over time [11; 7, 28; 25, 26, 38].

In this context, the poverty and social exclusion is treated as a multivariate phenomenon, which should be analyzed with application of multiple criteria decision analysis (MCDA) tools. In this case TOSPIS method was applied in order to obtain TMR of the phenomenon. In the method a choice of metric applied to calculation of the distance of objects from positive and negative ideas solutions is a crucial factor. In the current research generalized distance measure GDM is used, as the metric can be applied for variables measured on a ratio scale, an interval scale, an ordinal scale or a nominal scale. As a result, from methodological perspective, the objective of the article is to present applicability of the measure GDM in TOPSIS method.

[1] Nicolaus Copernicus University, Department of Economics,ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: adam.balcerzak@umk.pl.
[2] Nicolaus Copernicus University, Department of Econometrics and Statistics, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: michal.pietrzak@umk.pl.

2 Methodology and data

To determine TMD values TOPSIS method is most commonly used [2, 4, 5, 6 14, 19, 29]. The procedure of TMD assessing is based on the the comparison of objects to the pattern and anti-paten of development, in the case of TOPSIS usually called as positive or negative ideal solutions. The values of TMD measured with TOPSIS method are normalized and range from zero to unity. The high values of the TMD indicate high level of development of the investigated phenomenon for the selected object. In order to compare the objects with positive and negative ideal solutions, the distance is determined based on a given metric. The choice of the metric is of crucial importance and it should be primarily dependent on the type of data held. The metric that can be used for variables from any measurement scale is the generalized distance measure GDM proposed by Walesiak [35]. This metric can be used in the case of variables measured on the ratio scale, interval scale, ordinal scale or nominal scale [see 20, 36, 37].

The procedure of assessing TMD with application of TOPSIS based on Generalized Distance Measure GDM can describe in the following steps [37]:

1. The choice of diagnostic variables X_j (j=1,2,..,n), which describe a given economic phenomenon.
2. Establishing the set of economic objects O_i (i=1,2,...,m), for which the value of the TMD will be determined.
3. Determining the nature of diagnostic variables (stimulants or dis-stimulants).
4. Normalisation of the diagnostic variables, as a result a set of variables Z_j is obtained.
5. Establishing for diagnostic variables (stimulants) positive ideal solution P_j and negative ideal solution AP_j, which is given with equation 1. In the case of dis-stimulants the equations are applied the other way round. For dynamic research the values of positive and negative ideal solutions should be constant for the whole timespan of the research. This enables comparability of the research in different periods t.

$$P_j = \max_{it} z_{ijt} \quad , \quad AP_j = \min_{it} z_{ijt} \,. \tag{1}$$

6. Assessing the values of TMD_{it} with equation 2:

$$TMD_{it} = 1 - \frac{GDM_{it}^{P}}{GDM_{it}^{AP} + GDM_{it}^{P}}; \tag{2}$$

Where GDM_{it}^{P} is a distance of the object form the positive ideal solution, and GDM_{it}^{AP} is a distance of the object from negative ideal solution. Generalized distance measure GDM is given with equations 3 and 4 (see 36, 37]:

$$GDM_{it}^{P} = \frac{1}{2} - \frac{\sum_{j=1}^{m}(z_{ijt} - P_{kj})(P_{kj} - z_{ijt}) + \sum_{j=1}^{m}\sum_{\substack{l=1 \\ l\neq i,k}}^{n}(z_{ijt} - z_{ljt})(P_{kj} - z_{ljt})}{2\left[\sum_{j=1}^{m}\sum_{l=1}^{n}(z_{ijt} - z_{ljt})^2 \cdot \sum_{j=1}^{m}\sum_{l=1}^{n}(P_{kj} - z_{ljt})^2\right]^{\frac{1}{2}}}, \tag{3}$$

$$GDM_{it}^{AP} = \frac{1}{2} - \frac{\sum_{j=1}^{m}(z_{ijt} - AP_{kj})(AP_{kj} - z_{ijt}) + \sum_{j=1}^{m}\sum_{\substack{l=1 \\ l\neq i,k}}^{n}(z_{ijt} - z_{ljt})(AP_{kj} - z_{ljt})}{2\left[\sum_{j=1}^{m}\sum_{l=1}^{n}(z_{ijt} - z_{ljt})^2 \cdot \sum_{j=1}^{m}\sum_{l=1}^{n}(AP_{kj} - z_{ljt})^2\right]^{\frac{1}{2}}}, \tag{4}$$

where $i,l = 1,\ldots,n$ –number of the object, k – number of pattern of development and anti-pattern of development, $j = 1,\ldots,m$ – number of variable,

In the research the problem of poverty and social exclusion in Visegrad countries at regional level NUTS 1 was the subject of investigation. The phenomenon is the result of influence of many regional factors such as the situation on the local labour markets [24, 32,], actions of local authorities and quality of public services supporting sustainability of economy [1, 3, 15, 31] and others. These factors justify application of TOPSIS method for its measurement. Based on the aim of the article four diagnostic variables, which are suggested by Eurostat. The variables are given in table 1. The data was downloaded from Eurostat service: http://ec.europa.eu/eurostat/data/database.

Economic development	
X_1 – At-risk-of-poverty rate (Percentage of total population)	stimulant
X_2 – People living in households with very low work intensity (Percentage of total population aged less than 60)	stimulant
X_3 – Severe material deprivation rate (Percentage of total population)	stimulant
X_4 – People living in households with very low work intensity (Percentage of total population)	stimulant

Table 1 Diagnostic variables

3 Results and discussion

According to the presented procedure of TMD measurement, after the choice of diagnostic variables, normalization of the variables was performed. Zero-unitarization method was used for this purpose [9, 23]. Then the constant positive and negative ideal solutions for both years of the research and the distances of the objects form the positive and negative ideal solutions were assessed. Finally, the TMD for the year 2011 and 2015 has been calculated according to formula 2.

The obtained values of TMD enabled to propose ranking of regions, starting with the once with the lowest level of poverty and social exclusion. The results for the years 2011 and 2015 are given in table 2. Additionally, with application of natural breaks method the regions were grouped into three classes. In the class 1 the regions with the highest level of poverty and social exclusion can be found. In the class 2 the regions with average level of the phenomenon are grouped. In the class 3 there are the regions with the lowest level of poverty and social exclusion. The final results of classification are given in table 2 and figure 1.

Region	2011			2015			Change
	TMD	Class	Rank	TMD	Class	Rank	
Czech Republic	0,087	3	1	0,100	3	1	13,93%
Slovakia	0,243	3	2	0,241	2	2	-0,64%
Poland: Region Poludniowy	0,316	2	4	0,290	2	3	-8,23%
Poland: Region Centralny	0,294	2	3	0,292	2	4	-0,65%
Poland: Region Poludniowo-Zachodni	0,409	2	8	0,316	2	5	-22,63%
Poland: Region Pólnocno-Zachodni	0,395	2	6	0,398	2	6	0,79%
Poland: Region Pólnocny	0,391	2	5	0,493	1	7	26,28%
Hungary: Dunántúl	0,538	1	9	0,516	1	8	-4,13%
Poland: Region Wschodni	0,572	1	10	0,518	1	9	-9,51%
Hungary: Közép-Magyarország	0,398	2	7	0,525	1	10	31,81%
Hungary: AlföldésÉszak	0,844	1	11	0,831	1	11	-1,53%

Table 2 Ranking and grouping of regions based on the level of poverty and social Exclusion

In the year 2011 in the class 3 grouping the NUTS 1 regions with the lowest level of poverty one could find Czech Republic and Slovakia. The countries are characterized with high stability of the most important macroeconomic indicators and the best labor market situation. In the class 2, 5 five Polish regions and one Hungarian region Közép-Magyarország were assigned. In class 1 with the highest level of poverty and social Exclusion one Polish region (Wschodni) and the two Hungarian regions (Közép-Magyarország and AlföldésÉszak) were grouped. These regions are characterized with the worst socio-economic situation in both countries.

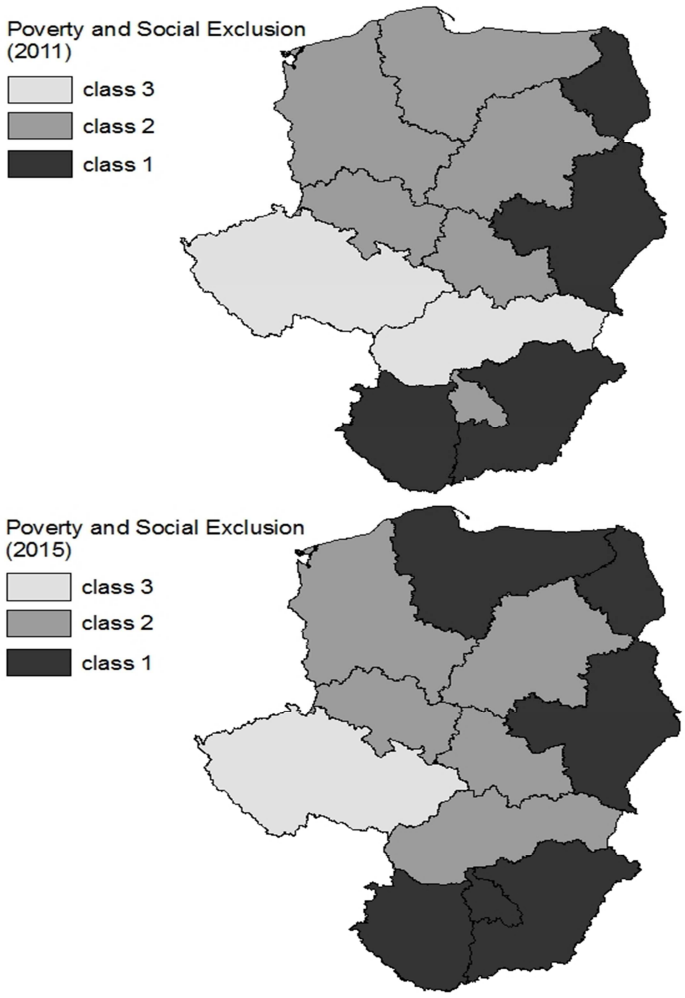


Figure 1 The level of Poverty and Social Exclusion in the year 2011 and 2015

Taking into account the percentage changes of TMD between 2011 and 2015, the relative situation of the five regions in this period has deteriorated. The relative poverty increased for the Czech Republic by 13.92%, for Region Północno-zachodni (Poland) by 0.79%, Region Północny (Poland) by 26.28%, and for Közép-Magyarország (Hungary) by 31.81%. On the other hand, in the case of seven regions, the situation has improved. The values of the measure decreased for Slovakia by 0.64%, for Region Południowy (Poland) by 8.23%, for the Region Centralny (Poland) by 0.65%, for the Region Południowo-Zachodni (Poland) by 22.63%, for Dunántúl (Hungary) by 0.13%, for Region Wschodni (Poland) by 9.51% and for AlföldésÉszak(Hungary) by 1.53%.

This dynamics has influenced the classification of the regions and their grouping in the year 2015. In spite of the negative tendency, Czech Republic was still classified as the NUTS 1 region with the lowest level of poverty and social Exclusion. Slovakia, was downgraded from class 3 in 2011 to class 2 in 2015. In class 1, the number of regions increased to five. Region Wschodni (Poland), Közép-Magyarország and AlföldésÉszak (Hungary) remained in class 1 in 2015. Region Północny (Poland) and Közép-Magyarország (Hungary) were downgraded in 2015 from class 2 to class 1. Four Polish regions: Region Poludniowy, Region Centralny, Region Poludniowo-Zachodni, Region Pólnocno-Zachodni, as in the year 2011, were grouped class 2 in the year 2015.

4 Conclusions

The aim of the research was to analyze the phenomenon of poverty and social exclusion in Visegrad countries at regional level (NUTS 1) in the years 2011 and2015. The poverty was treated as a multiple criteria phenomenon. Thus, multiple-criteria decision analysis tool (TOPSIS method) was used. In comparison to the standard TOPSIS method, in the current research generalized distance measure GDM was used for calculation the distance of objects from positive and negative ideas solutions. The research confirms universality of the distance measure, which can be used for variables measured on the ratio scale, the interval scale, the ordinal scale or the nominal scale.

The conducted classification of the object confirms significant diversity between the analyzed regions. The level of poverty and social exclusion is the result of many factors, which from the national and regional policy

perspective impede the effectiveness of potential actions of authorities. This can be seen in the obtained results, where in the years 2011-2015 one cannot see a specific tendency for all the regions under research. Just the opposite the changes of the value of TMD for the analyzed regions vary from 1 to 30%. The obtained results also confirm that the phenomenon is characterized with a high degree of inertia, as the changes of the relative positions of the regions in the ranking between the analyzed years are not big.

References

[1] Adamowicz, M., and Machla, A.: Small and medium enterprises and the support policy of local government. *Oeconomia Copernicana*, **7(3)** (2016), 405-437.

[2] Ali, R., Lee, S., and Chung, T.: Accurate multi-criteria decision making methodology for recommending machine learning algorithm. Expert Systems with Applications, **71** (2017), 257-278.

[3] Ambroziak, A. A.: Income tax exemption as a regional state aid in special economic zones and its impact upon development of Polish districts. *Oeconomia Copernicana*, **7(2)** (2016), 245-267.

[4] An, D., Xi, B., Ren, J. Z., Wang, Y., Jia, X., He, C., and Li, Z.: Sustainability assessment of groundwater remediation technologies based on multi-criteria decision making method. *Resources Conservation and Recycling*, **119** (2017), 36-46.

[5] Arikan, F., and Citak, S.: Multiple criteria inventory classification in an electronics firm. *International Journal of Information Technology & Decision Making*, **16(2)** (2017), 315-331.

[6] Baykasoglu, A., and Golcuk, I.: Development of an interval type-2 fuzzy sets based hierarchical MADM model by combining DEMATEL and TOPSIS. *Expert Systems with Applications*, **70** (2017), 37-51.

[7] Balcerzak, A. P.: Effectiveness of the institutional system related to the potential of the knowledge based economy. *Ekonomista*, **6** (2016), 711-739.

[8] Balcerzak, A. P.: Europe 2020 strategy and structural diversity between old and new member states. Application of zero-unitarizatin method for dynamic analysis in the years 2004-2013.*Economics & Sociology*, **8(2)** (2015), 190-210.

[9] Balcerzak, A. P.: Fiscal burden in the European Union countries. *Economic Annals XXI*, **161(9-10)** (2016), 4-6.

[10] Balcerzak, A. P., and Rogalska, E.: Non-keynesian effects of fiscal consolidations in Central Europe in the years 2000-2013.In: *Entrepreneurship, business and economics - Vol. 2. Proceedings of the 15th Eurasia Business and Economics Society* (Bilgin, M. H., and Danis, H., eds.). Springer International Publishing,2016, 271-282.

[11] Balcerzak, A. P.: Technological potential of European Economy. Proposition of measurement with application of multiple criteria decision analysis. *Montenegrin Journal of Economics*, **12(3)** (2016), 7-17.

[12] Balcerzak, A. P.: Multiple-criteria evaluation of quality of human capital in the European Union countries.*Economics & Sociology*, **9(2)** (2016), 11-27.

[13] Balcerzak, A. P., and Pietrzak, M.B.:Quality of institutions for knowledge-based economy within new institutional economics framework. Multiple criteria decision analysis for European countries in the years 2000–2013. *Economics & Sociology*, **9(4)**(2016), 66-81.

[14] Balcerzak, A. P., and Pietrzak, M. P.: Application of TOPSIS method for analysis of sustainable development in European Union countries. In: *The 10th international days of statistics and economics. Conference proceedings. September 8-10, 2016* (Loster T., and Pavelka, T. eds.). Libuse Macakova, Melandrium, Prague, 82-92.

[15] Czaplak, J.: Analysis of the effectiveness of public services provision by local governments in the years 2007–2013. *Oeconomia Copernicana*, **7(2)** (2016), 223-243.

[16] Furková, A., and Chocholatá, M.: Interregional R&D spillovers and regional convergence: a spatial econometric evidence from the EU regions. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, **12(1)** (2017), 9-24.

[17] Hadaś-Dyduch, M., Balcerzak, A. P., and Pietrzak, M.B.: Wavelet analysis of unemployment rate in Visegrad countries. In:*16th International Scientific Conference Globalization and Its Socio-Economic Consequences. University of Zilina, The Faculty of Operation and Economics of Transport and Communication, Department of Economics, 5th – 6th October 2016. (Part II.)* (Kliestik, T. ed.). University of Zilina, Zilina, 2016, 595-602.

[18] Hellwig, Z.: Procedure of evaluating high-level manpower data and typology of countries by means of the taxonomic method. In: *Towards a system of human resources indicators for less developed countries, Papers prepared for a UNESCO research project* (Gostowski, Z. ed.).Ossolineum-The Polish Academy of Sciences, Wrocław, 1972, 115-134.

[19] Hwang, C. L., and Yoon, K.: *Multiple attribute decision making: methods and applications*. Springer, Heidelberg, 1981.

[20] Jajuga, K., Walesiak, M., and Bąk, A.: On the general distance measure. In: *Exploratory data analysis in empirical research* (Schwaiger, M. and Opitz, O., eds.). Springer-Verlag, Berlin, Heidelberg, 104-109.
[21] Jantoń-Drozdowska, E., & Majewska, M.: Investment attractiveness of Central and Eastern European countries in the light of new locational advantages development. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, **11(1)** (2016), 97-119
[22] Kuc, M.: Social convergence in Nordic countries at regional level. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, **12(1)** (2017), 25-41.
[23] Kukuła, K., and Bogocz, D.: Zero unitarization method and its application in ranking research in agriculture. *Economic and Regional Studies*, **7(3)** (2014), 5-13.
[24] Łaszkiewicz, E.: Determinants of hourly wages inequality in selected European metropolises. The results from the multilevel modelling. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, **11(4)** (2016), 853-869.
[25] Łyszczarz, B.: Public-private mix and performance of health care systems in CEE and CIS countries. *Oeconomia Copernicana*, **7(2)**(2016), 169-185.
[26] Małkowska, A., & Głuszak, M.: Pro-investment local policies in the area of real estate economics - similarities and differences in the strategies used by communes. *Oeconomia Copernicana*, **7(2)**, (2016), 269-283.
[27] Müller-Frączek, I., and Pietrzak, M. B.: Spatial analysis of the unemployment rate in Poland in 2004-2008. In: *Hradec Economic Days 2011* (Jedlicka, P., ed.). Gaudeamus, Hradec Králové, 2011, 205-209.
[28] Pietrzak, M. B.: The problem of the inclusion of spatial dependence within the TOPSIS Method. *Montenegrin Journal of Economics*, **12(3)** (2016), 69-86.
[29] Pietrzak, M. B., and Balcerzak, A. P.: Quality of human capital and total factor productivity in new EU member states. In: *The 10th international days of statistics and economics. Conference proceedings. September 8-10, 2016* (Loster, T. and Pavelka, T. eds.). Libuse Macakova, Melandrium, Prague, 1492-1501.
[30] Pietrzak, M. B., and Balcerzak, A. P.: Assessment of socio-economic sustainability in new European Union members states in the years 2004-2012.In: *The 10th professor Aleksander Zelias international conference on modelling and forecasting of socio-economic phenomena*. *Conference proceedings* (Papież, M., and Śmiech, S, eds.). Foundation of the Cracow University of Economics, Cracow, 2016, 120-129.
[31] Pietrzak, M. B., and Balcerzak, A. P.: Socio-economic sustainability in Poland. SEM analysis at regional level. In:*16th International scientific conference globalization and its socio-economic consequences. University of Zilina, The Faculty of Operation and Economics of Transport and Communication, Department of Economics, 5th – 6th October 2016. (Part IV.)* (Kliestik, T. ed.). University of Zilina, Zilina, 2016, 1704-1712.
[32] Pietrzak, M. B., and Balcerzak, A. P.: A spatial SAR model in evaluating influence of entrepreneurship and investments on unemployment in Poland. In: *Proceedings of the international scientific conference quantitative methods in economics multiple criteria decision making XVIII* (Reiff, M., and. Gezik, P. eds.). Letra Interactive, Vratna, 2016, 303-308.
[33] Pietrzak, M. B., Wilk, J., Kossowski, T., and Bivand, R. (2014). The identification of spatial dependence in the analysis of regional economic development - Join-count test application. In: *Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena* (Papież, M., and Śmiech, S. eds.). Foundation of the Cracow University of Economics, Cracow, 135-144.
[34] Pietrzak, M. B, Wilk, J., and Siekaniec, M. (2013). The impact of metropolitan areas on internal migrations in Poland. The case of southern regions. In: *Proceedings of the 7th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena* (Papież, M., and Śmiech, S., eds.). Foundation of the Cracow University of Economics, Cracow, 124-132.
[35] Walesiak, M.: Distance measure for ordinal data. *Argumenta Oeconomica*, **2(8)** (1999), 167-173.
[36] Walesiak, M.: *Uogolniona miara odleglosci w statystycznej analizie wielowymiarowej*. Wydawnictwo AE, Wrocław, 2006.
[37] Walesiak, M.: *Uogólniona miara odległosci GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*. Wydawnictwo UE, Wrocław, 2016.
[38] Żelazny R., & Pietrucha J.: Measuring innovation and institution: the creative economy index. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, **12(1)** (2017), 43-62.

Modification of EVM by scenarios

Jan Bartoška[1], Tomáš Šubrt[2], Petr Kučera[3]

Abstract. The paper brings a new perspective to the monitoring of actual costs for Earned Value Management in projects. Costs in phase of implementation of the project vary depending on the internal and external factors, which cannot be clearly predicted in advance. In the practice of project management, actual costs (AC) are usually not sufficiently known and so they are very often only estimated during the project duration. For the Earned Value Management (EVM), while tracking project progress, knowledge of the AC is crucial. Any distortion or incorrect estimate of the AC leads to distortion of all EVM characteristics. CPI (Cost Performance Index) and CV% (Cost Variance) within EVM describe the status and progress of the project and demonstrate its success, are based on correct value of AC, and thus AC bias leads to bias and distortion of any prediction of future cost development (ETC, EAC, etc.) . Our proposal published in this paper is based on the use of optimistic, neutral and pessimistic scenarios for the development of current costs during project implementation. Proposals of scenarios are based on mathematical model designed by authors and complementing already existing models for the planned value (PV) estimation within EVM (published by authors in previous papers). This article aims to provide a comprehensive proposal how to modify the EVM based on mathematical models for AC and PV estimations.

Keywords: Project Management, Earned Value Management, Work Effort, Actual Cost, Cone of Uncertainty, Optimistic and Neutral and Pessimistic Scenario.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Earned Value Management [15, 13 or 6] is based on comparing the Baseline and Actual plan of the project realization. Baseline and Actual plan is determined by the partial work of single resources in the particular project activities. Baseline is always based on expected resource work contours and the impact of human agent is usually not included. The impact of human agent is usually expected only in the actual course of the project. The versatility of the human agent in projects can be described also by the "*Parkinson's first law*" [12]. It is natural for people to distribute work effort irregularly to the whole time sequence which was determined by the deadline of the task completion. The questions of "Parkinson's first law" in project management are further dealt with in e.g. [5].

Work effort of an allocated resource has very often been researched in projects from the area of information technologies and software development, as these projects contain a high level of indefiniteness, and even common and routine tasks are unique. At the same time, it concerns the area where it is possible to find a great number of approaches to estimate how work-intensive the project will be or how long the tasks will take, and also case studies. The proposal for mathematical apparatus for planning the course of tasks within a case study is dealt with for instance in [11 or 1]. The authors Özdamar and Alanya [11] propose a particular pseudo-heuristic approach to estimate the tasks course where the indefiniteness in the project is expressed by fuzzy sets. Barry et al. [1] concentrate on the existence and expression of the relation between project duration and total effort and in their theoretical starting points they emphasize the dynamics of the relation between the effort and project duration when a self-strengthening loop can be expected. The work effort can be described also using system dynamic models as presented e.g. in a study from project management teaching by Svirakova [14] or in a study from project simulation model by Lacko [8]. The others who research the project complexity and work effort are for instance Clift and Vandenbosh [4], who point out a connection between the length of life cycle and project management structure where a key factor is again a human agent.

Among properties typical for projects belongs their variability during their realization. Although there is an effort to perform repeatedly corrections based on new budget and time term estimates during the project realization, project managers still work with a high uncertainty rate [4, 8, 9, 10 or 13]. The uncertainty of the estimates,

[1] Department of Systems Engineering FEM CULS, Kamýcká 129, Prague, bartoska@pef.czu.cz.
[2] Department of Systems Engineering FEM CULS, Kamýcká 129, Prague, subrt@pef.czu.cz.
[3] Department of Systems Engineering FEM CULS, Kamýcká 129, Prague, kucera@pef.czu.cz.

i.e. their variability, decreases during the project, as proven in NASA study [10] or, as well, specified by Little [9]. It is possible to use this decreasing variability of the budget estimates as a basis for a new modification of EVM, in particular, the Actual Costs (AC). The aim of this paper is to modify EVM based on a proposal of the mathematical model for Cone of Uncertainty with an extension into an optimistic, neutral and pessimistic scenario of the project costs development. The paper builds on the previous authors' works [2, 3 or 7].

2 Materials and methods

2.1 Student Syndrome phenomenon

If there is a deadline determined for the completion of a task and a resource is a human agent, the resource makes its effort during the activity realization unevenly and with a variable intensity. Delay during activity realization with human resource participation leads to stress or to tension aimed at the resource or the tension of the resource him/herself. The development and growth of the tension evokes the increase in the work effort of the human agent allocated as a resource. More detailed approach can be found in [2].

2.2 Mathematical model of the Student Syndrome

Authors in previous paper [2] propose a mathematical expression of the Student Syndrome. Its brief description follows: First, a function expressing the proper Student Syndrome denoted by p_1 is introduced. It has three minima $p_1(t) = 0$ in $t = 0$, $t = 0.5$, and $t = 1$; and two maxima: former one close to the begin and latter one close to the end of the task realization. Beside this, functions denoted by p_2 expressing the resource allocation according to single standard work contours of flat, back loaded, front loaded, double peak, bell and turtle are proposed. All these functions are in the form of 4th degree polynomial. To express the strength of the Student Syndrome manifestation during the realization of a task the rate r of the Student Syndrome is introduced. It acquires values between 0 and 1 ($r = 0$ represents a situation when the Student Syndrome does not occur at all and the resource keep the work contour exactly; $r = 1$ means that the Student Syndrome manifests in its all strength and the resource absolutely ignore the work contour). As a result, the resource work effort during a real task realization can be modeled using function $p = rp_1+(1–r)p_2$. More detailed approach can be found in [2].

2.3 Modification of the Planned Value by Work Effort

The EVM extension in the form of Planned Value (*PV* or *BCWS*) parameter modification for different work contours (turtle, bell, double peak, back loaded, front loaded, etc.) which is described below and applied in a case study is based on previous work of the authors of this paper [2, 7].

This approach can be applied when computing the *PV* of an activity in the project. It is computed in the classical way using the formula:

$$PV = \%PC \,.\, BAC \tag{1}$$

where *%PC* is the percentage of the work planned by the work calendar (planned completion percentage) and *BAC* is Budget at Completion of the project. The share of the whole work effort as a part of task duration requires can be calculated for a single resource as:

$$\int_0^a p(t)dt = 1 \tag{2}$$

Let there are n resources, indexed by 1, 2, …, n, allocated at the task. Let r_k, p_{1k}, p_{2k} denotes r, p_1, p_2 for k-th resource. Then the *PV* can be calculated (where t is a time moment in the project):

$$PV_a = \left(\sum_{k=1}^{n} \int_0^a (r_k\, p_{1k}(t) + (1 - r_k)\, p_{2k}(t))dt \right) \cdot BAC \tag{3}$$

The resource work effort affects the growth of *PV*. It is not possible to expect uniform increase of *PV* in general and in case of all project tasks. In case of changing *PV*, EVM may provide fundamentally different results for the assessment of the state and development of the project.

2.4 Linear and Non-Linear PV

The work [3] discusses an application of a newly proposed *PV* (*BCWS*) calculation for monitoring the project duration within EVM. As the case study in [3] shows, as far as the task nature (work contour) and the human resource impact (Student Syndrome phenomenon) is comprised, the planned course of work effort (Non-Linear *PV*) may significantly differ from the commonly expected one (Linear *PV*). A possible decline or increase of resources work effort, which is not evident in case of the uniform work plan with uniform work effort, may manifest itself in an irregular increase or decrease of the Earned Value. This may result in malfunctioning EVM and project failure. For more details see previous paper [3].

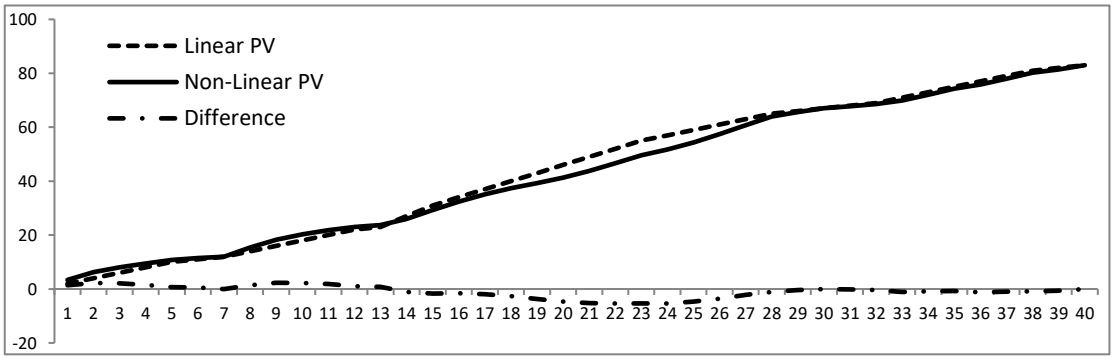


Figure 1 Case Study – difference between Linear and Non-Linear BCWS [3]

The difference in course between the Linear and Non-Linear *PV* leads within EVM to different conclusions in the assessment of the project status and development [3]. Overmuch positive or negative project assessment may result in uneconomic decisions with fatal consequences. With the inclusion of the human factor impact and various work contours and the derivation of the Non-Linear *PV* it is possible to obtain a much more accurate (though possibly less pleasant) illustration of the project.

2.5 Cone of Uncertainty

The Cone of Uncertainty (CU) is a phenomenon proven in NASA studies [10] and often cited in professional literature, e.g. by Little [9]. CU manifests itself in projects during their realization in the form of decreasing budget and time estimates.

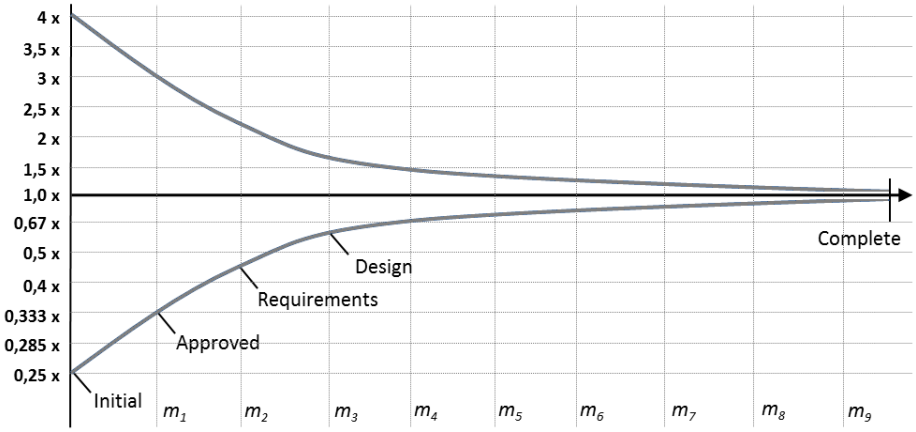


Figure 2 Cone of Uncertainty with milestones (m_i) [10 or 9]

In case of the budget we can expect up to four times bigger estimate inaccuracy at the beginning of the project, in case of the time terms up to two times bigger estimate inaccuracy [10]. For the purposes of this paper we will focus on the estimate variability for the project budget.

3 Results and Discussion

First, to enable to propose an extension of the new approach to EVM according to [2, 3 or 7], it is necessary to propose a mathematical model for CU. Subsequently, the newly proposed model for CU will be applied to create an optimistic, neutral and pessimistic scenario of the Actual Costs (AC) development.

3.1 Mathematical model of Cone of Uncertainty

The form and course of CU can be expressed in the form of the following mathematical model:

$$c_t^1 = 1 + 3e^{-4(t/T)} \tag{4}$$

where t is the serial number of the current day (month) of the project duration and T is the total number of the days (months) of the project duration.

Formula (4) enables to quantify the optimistic estimate deviation in any time moment during the project, i.e. it explains decreasing and thus improving variability of costs expended within the project. This mathematical model can be modified for pessimistic estimate development in the CU form in the following way:

$$c_t^2 = (4 - 3e^{-4(t/T)}) / 4 \tag{5}$$

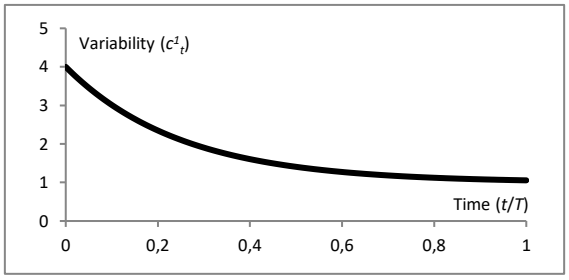


Figure 3 Optimistic development (c^1_t) – estimates get better

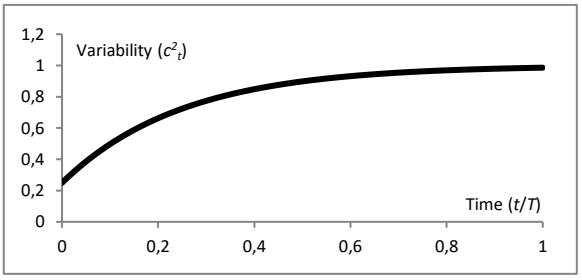


Figure 4 Pessimistic development (c^2_t) – estimates worsen

3.2 Optimistic, neutral and pessimistic scenario in projects

Budget spending scenario may be optimistic, neutral or pessimistic. The costs currently spent are not always during the project realization sufficiently known, and, to capture the state and development of the project, even in the short term, it is necessary to perform an approximate estimation of the actually spent costs based on planned course of work. In case of the optimistic scenario, AC may be at the beginning of the project higher than the reality and during the project they may decrease, get better – the project management is pleasantly surprised. In case of the pessimistic scenario, it is possible to assume the contrary, a low cost estimation at the beginning and high to the end of the project realization – the project management is surprised unpleasantly. The neutral scenario may be expressed by the balanced weighted sum of the optimistic and pessimistic scenario. To express mathematically all the scenarios, we can apply the mathematical model for CU (formulas (4) and (5)) extended by an optimistic index form the interval <0, 1>:

$$^{CU}AC_t^i = i.PV_t / c_t^1 + (1-i).PV_t / c_t^2 \tag{6}$$

In formula (6) we propose the calculation of AC based on a possible scenario and including CU impact. In the same time, the relation between estimation variability in time (c^1_t or c^2_t) and the planned work course (PV_t) is given in the form of ratio. As a result, we obtain an estimation of actual costs including uncertainty.

3.3 Case Study

The application of proposals (4), (5) a (6) can be demonstrated on the case study from the previous work [3]. The original case was extended by AC calculations for optimistic (i = 0.9), neutral (i = 0.5) and pessimistic (i = 0.1) scenario including Measured AC:

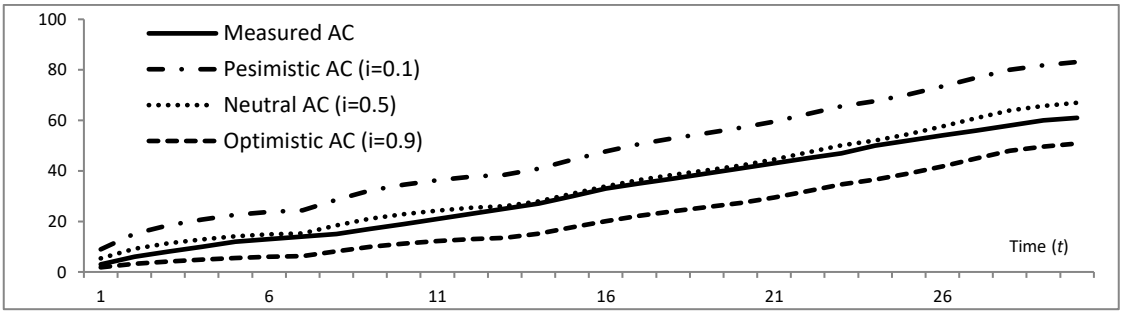


Figure 5 Case Study – AC calculation for single AC scenarios

The considered extension of the case from [3] can be expanded by the calculation of the Estimate at Completion (EAC), i.e. by the calculation of the estimate of costs for the project completion. Again, all the outlined scenarios are applied:

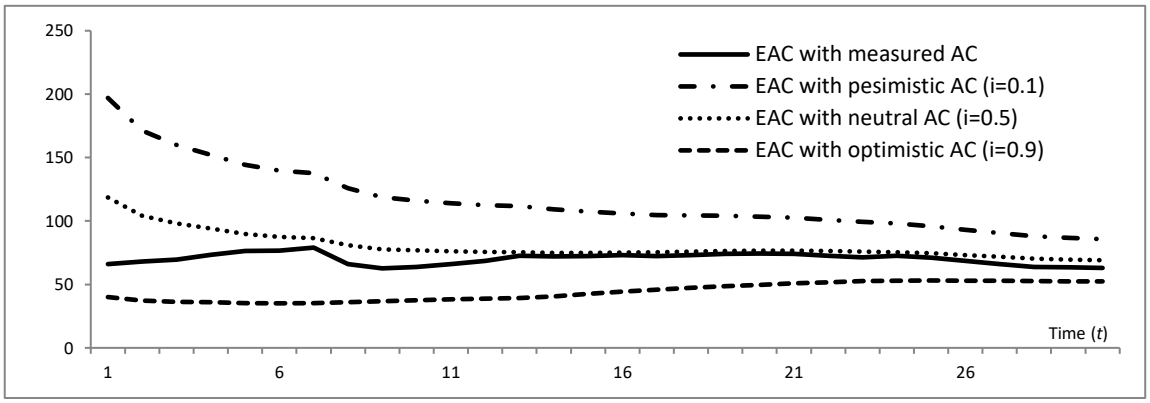


Figure 6 Case Study – Calculation of the costs estimate for the project completion (EAC) according to the single AC scenarios

From Figures 5 and 6 it is obvious that the neutral scenario, given by the balanced weighted sum of the optimistic and pessimistic development of AC estimates according to CU approaches to actual, later found out, Measured AC. Whereas the model for the optimistic and pessimistic scenario marks out borders for possible estimates of spending the budget – the variability and the impact of uncertainty in the estimates is apparent.

Moreover, as long as we involve the fact that all the applied models also express (mathematically implicitly include) the human factor impact (in the form of the Student Syndrome phenomenon and Parkinson's project law, see previous authors' works [2, 3 and 7]), the resulting EMV modification affects in a complex way the uncertainty of projects, and, beside this, in a new unique manner.

4 Conclusion

The paper quantifies in an original way the CU phenomenon and concludes an extensive and complex modification of EVM for project management – the paper thus builds on previous works [2, 3 or 7]. A part of the paper is a proper proposal of mathematical models for CU and its application in the calculation of partial AC values during the project realization. Beside this, mathematical models are used to propose the optimistic, neutral and pessimistic scenario of the project costs development. In the case study in which we continue from the previous paper [3] we present single scenario courses and face them up to actually found out costs. It is obvious from the paper results that it is possible, based on the mathematical model for CU, to model borders for estimates of costs expended within the project. This particular benefit can be utilized in the eventual practice, especially in case that a moment without any unequivocal possibility to trace or verify the Actual Costs happens in a project and it is a need to estimate the spent budget based on the planned work of resources.

The mathematical model proposals in the paper complete and conclude the EVM modification proposed in authors' works [2, 3 or 7]. All three fundamental input parameters (PV, EV, AC) are thus modified into the form of proper mathematical models including either the human factor impact (Student Syndrome phenomenon, Parkinson's first law) or uncertainty impact (CU). As a result, the development of the project, its costs and extent fulfillment, can be modeled in the closest connection to the real environment of organizations.

References

[1] Barry, E., Mukhopadhyay, T., and Slaughter, A. S. 1: Software project duration and effort: An empirical study. *Information Technology and Management* **3** (2002), 113–136.

[2] Bartoška, J., Kučera, P., and Šubrt, T.: Modification of the BCWS by Work Effort. In: *Proceedings of the 33rd International Conference Mathematical Methods in Economics MME 2015* (Martinčík, D., Ircingová, J., and Janeček, P., eds.). University of West Bohemia, Plzen, 2015, 19-24.

[3] Bartoška, J., Kučera, P., Šubrt, T.: Modification of the EVM by Work Effort and Student Syndrome phenomenon. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics*. Technical University Liberec, 2016, pp. 25-30. ISBN 978-80-7494-296-9.

[4] Clift, T., and Vandenbosh, M.: Project complexity and efforts to reduce product development cycle time. *Journal of Business Research* **45** (1999), 187–198.

[5] Gutierrez, G.J., and Kouvelis, P.: Parkinson's law and its implications for project management. *Management Science* **37** (8) (1991), 990-1001.

[6] Kim, B., C., and Kim, H., J.: Sensitivity of earned value schedule forecasting to S-curve patterns. *Journal of Construction Engineering and Management* **140** (7) (2014), 14-23.

[7] Kučera, P., Bartoška, J., and Šubrt, T.: Mathematical models of the work contour in project management. In: *Proceedings of the 32nd International conference on Mathematical Methods in Economics* (Talašová, J., Stoklasa, J., and Talášek, T., eds.). Palacký University, Olomouc, 2014, 518-523.

[8] Lacko, B.: Modelování a simulace v oblasti projektového řízení. In: *Sborník semináře MSP 2001 - Modelování a simulace projektů*. VUT, Brno, 2001, str. 5-20.

[9] Little, T.: Schedule estimation and uncertainty surrounding the cone of uncertainty. *IEEE Software* **23** (3) (2006), 48-54.

[10] NASA. *Manager's Handbook for Software Development, Revision 1*. Document number SEL-84-101. Greenbelt, Maryland: Goddard Space Flight Center, 1990.

[11] Özdamar, L., and Alanya, E.: Uncertainty modelling in software development projects (with case study). *Annals of Operations Research* **10** (2001), 157-178.

[12] Parkinson, C.N.: *Parkinson's Law and Other Selected Writings On Management*. 1st ed., Federal Publications (S) Pte Ltd, Singapure, 1991.

[13] Project Management Institute. *A Guide to the Project management body of knowledge (PMBOK® Guide).* Fifth edition. Newtown Square: Project Management Institute Inc., p. 589, 2013.

[14] Svirakova, E.: System dynamics methodology: application in project management education. In: *Proceedings of the 11th International Conference on Efficiency and Responsibility in Education* (Houška, M., Krejčí, I., and Flégl, M., eds.). Czech University of Life Sciences, Prague, 2014, 813-822.

[15] Vandevoorde, S., and Vanhoucke, M.: A comparison of different project duration forecasting methods using earned value metrics. *International Journal of Project Management* **24** (4) (2005), 289–302.

DEA Models in Evaluation of Factors of Temporary Absence from Work in the Czech Republic

Markéta Bartůňková[1], Jan Öhm[2], Jiří Rozkovec[3]

Abstract. The length of temporary absence from work is a variable which depends on many factors. Just as the absence itself. From another point of view it has an impact on economics, public health care etc. The Czech Republic although it is not a large country shows significant differencies among regions concerning the working incapability. In connection with that this paper focuses on economic, ecological and demographical factors and via models of data envelopment analysis (DEA) compares Czech regions from this point of view. The first part of the paper analysis the correlations of selected variables. The next part describes the research methodology. BCC-I and BCC-O models were used with regard to the involved factors. The final part includes the summary of evaluated efficiencies and comparison of the regions. All the data was provided by the Institute of Health Information and Statistics of the Czech Republic and the Czech Statistical Office.

Keywords: temporary absence from work, data envelopment analysis, linear models, working incapability.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Incapacity for work is one of the social risks which accompanies society from the beginning of modern era (and had existed definitely also in traditional society although it had not been qualified as social). Thus many authors (e.g. Esping-Andersen, Bonoli, etc. - see [2], [4]) consider this risk as the old social risk. At the beginnings of social security system governments struggled especially with working incapability. The first complex solution of this risk was established in Germany where several law acts were introduced between 1883 and 1889. Before this reform the "occupational accident law" had been accepted in 1871. In present days is common that work absenteeism is covered by insurance or tax scheme and the risk costs are in a certain way diversified among government, employer and employee. The legitimate question is how much and on which factors temporary work absence depends. Definitely, there is also a reverse macroeconomic impact of work incapacity on GDP and other economic indicators.

This paper deals with a comparison of regions of the Czech Republic (follows as ČR) from the point of view of a temporary absence from work. In ČRthere is about 66 % of population between 15-64 years, i. e. people with economic activity. During period 2010-2015 more than 8 milions records of temporary absence from work was registered where its average length by regions varies from 22 to 60 days, it means from 1 to 3 months. Such a long absence has a negative economic impact on public health care expenses, GDP etc. To reduce such unfavourable effects the state invests, for instance, to medical devices or to the protection of environment. Here we try to evaluate and compare the efficiency of such investments in Czech regions in connection with a differencies in economic and demographical development etc.

2 Methodology

For the evaluation of efficiency we use the data envelopment analysis (DEA), specifically BCC-I (input-oriented) and BCC-O (output-oriented) models (see [3], [5]). As the inputs we used:

- number of medical devices (as NMD),
- volume of investments into environment (as IIE).

The outputs are:

[1]Technical University of Liberec, Faculty of Economics, department of economic statistics, Voronezska 13, 46001 LIBEREC, marketa.bartunkova@tul.cz
[2]Technical University of Liberec, Faculty of Economics, department of economic statistics, Voronezska 13, 46001 LIBEREC, jan.ohm@tul.cz
[3]Technical University of Liberec, Faculty of Economics, department of economic statistics, Voronezska 13, 46001 LIBEREC, jiri.rozkovec@tul.cz

- GDP per capita,
- life expectancy (as LE),
- proportion of time of presence in the work as a complement of absence from the work (as TP).

The first step is not the evaluation of efficiency, but a correlation analysis of used inputs and outputs. Because DEA models mentioned above are linear, it is necessary to test the correlation of selected inputs, outputs respectively, to eliminate dependent variables from the models (more in [6], [7]). Then the DEA models will be used for evaluation of efficiency using *Solver* in *MS Excel 2007*.

Remark 1. For simplification there are further used letters instead of long names of Czech regions. The letters correspond to the car registrations in ČR- see Table 1.

Region	Praha	Jihomoravský	Jihočeský	Pardubický	Královehradecký
Letter	A	B	C	E	H
Region	Vysočina	Karlovarský	Liberecký	Olomoucký	Plzeňský
Letter	J	K	L	M	P
Region	Středočeský	Moravskoslezský	Ústecký	Zlínský	
Letter	S	T	U	Z	

Table 1 Symbols of Czech Regions

2.1 Correlation of Variables

The test of correlation coefficient we can find for instance in [1]. Here just let us denote that we check the correlation of all pairs of outputs for all 14 regions in every year from 2010-2015 (there are three pairs, if there are three outputs) and in the same way all pairs of inputs (this is only one, if we have only two inputs) - see Section 2. All statistical calculations were done in *STATGRAPHICS Centurion XVII* at significance level $\alpha = 0,05$.

Year	*r*(GDP, LE)	*r*(GDP, TP)	*r*(LE, TP)
2010	0,498	-0,048	-0,030
2011	0,478	-0,056	0,109
2012	0,471	-0,005	-0,041
2013	0,564	-0,035	-0,113
2014	0,501	0,116	-0,001
2015	0,500	0,216	-0,017

Table 2 Outputs Correlation

From the test statistic

$$T = \frac{r(X,Y)}{\sqrt{1-r^2(X,Y)}}\sqrt{n-2} \sim t(n-2) \tag{1}$$

and the rejection area $W = \left\{T : |T| \geq t_{1-\frac{\alpha}{2}}(n-2)\right\}$ for $n = 14$ we get that the null hypothesis of independence for a particular pair of outputs is accepted when $|r(X,Y)| \leq 0,57$. So for all tested pairs this hypothesis was accepted, i. e. they are independent and all tested pairs of outputs can be used in the model together.

Year	2010	2011	2012	2013	2014	2015
r(NMD, IIE)	0,851	0,812	0,893	0,841	0,841	0,757

Table 3 Inputs Correlation

Contrary the inputs are strongly correlated according the results in Table 3 as the correlations are greater than 0,57. Therefore it is necessary to reduce their number and for the DEA models choose just one of them. In this paper we decided to use only number of medical devices (NMD) as the input. All the data used in this paper was provided by the Institute of Health Information and Statistics of the Czech Republic and the Czech Statistical Office.

2.2 DEA models

For the evaluation of efficiency we chose BCC model created by Banker, Charnes and Cooper in 1984. As usual the inputs are denoted as matrix $\mathbf{X}_{r\times n}, r = 1, n = 14$, outputs as matrix $\mathbf{Y}_{s\times n}, s = 3, n = 14$. The output-oriented model BCC-O looks like this:

$$\eta(\eta_B, \vec{\lambda}) = Max\, \eta_B \tag{2}$$

subject to

$$\begin{aligned} \mathbf{X}\vec{\lambda} &\leq \vec{x_r} \\ \eta_B \vec{y_r} - \mathbf{Y}\vec{\lambda} &\leq \vec{o} \\ \vec{e_n}^T \vec{\lambda} &= 1 \\ \vec{\lambda} &\geq \vec{o} \end{aligned}$$

The intput-oriented model BCC-I:

$$\theta(\theta_B, \vec{\lambda}) = Min\, \theta_B \tag{3}$$

subject to

$$\begin{aligned} \theta_B \vec{x_r} - \mathbf{X}\vec{\lambda} &\leq \vec{o} \\ \mathbf{Y}\vec{\lambda} &\leq \vec{y}_r \\ \vec{e_n}^T \vec{\lambda} &= 1 \\ \vec{\lambda} &\geq \vec{o} \end{aligned}$$

These models are the first phase of optimization. For each decision making unit U_r, i. e. r^{th} region, we get $\eta_B^\star$, $\theta_B^\star$ respectively. Then the following models are being solved:

BCC-O Model

$$\omega(\vec{\lambda}, \vec{s}_+, \vec{s}_-) = Min\, \vec{e_s}^T \vec{s}_+ + \vec{e_r}^T \vec{s}_- \tag{4}$$

subject to

$$\begin{aligned} \vec{s}_- &= \vec{x_r} - \mathbf{X}\vec{\lambda} \\ \vec{s}_+ &= \mathbf{Y}\vec{\lambda} - \eta_B^\star \vec{y_r} \\ \vec{e_n}^T \vec{\lambda} &= 1 \\ \vec{s}_+ &\geq \vec{o} \\ \vec{s}_- &\geq \vec{o} \\ \vec{\lambda} &\geq \vec{o} \end{aligned}$$

BCC-I Model

$$\omega(\vec{\lambda}, \vec{s}_+, \vec{s}_-) = Max\, \vec{e_s}^T \vec{s}_+ + \vec{e_r}^T \vec{s}_- \tag{5}$$

subject to

$$\begin{aligned} \vec{s}_- &= \theta_B^\star \vec{x_r} - \mathbf{X}\vec{\lambda} \\ \vec{s}_+ &= \mathbf{Y}\vec{\lambda} - \vec{y}_r \\ \vec{e_n}^T \vec{\lambda} &= 1 \\ \vec{s}_+ &\geq \vec{o} \\ \vec{s}_- &\geq \vec{o} \\ \vec{\lambda} &\geq \vec{o} \end{aligned}$$

where $\vec{e}_s = (1, 1, \ldots, 1) \in \mathbb{R}^s$, $\vec{e}_r = (1, 1, \ldots, 1) \in \mathbb{R}^r$, $\vec{e}_n = (1, 1, \ldots, 1) \in \mathbb{R}^n$ respectively. Vectors $\vec{s}_+$ and $\vec{s}_-$ are slacks of outputs or surpluses of inputs. The optimal solution then involves $(\theta_B^\star, \vec{\lambda}^\star, \vec{s}_+^\star, \vec{s}_-^\star)$. If $\theta_B^\star = 1$ and $\vec{s}_+^\star = \vec{o}$ and $\vec{s}_-^\star = \vec{o}$, then the decision making unit is *BCC-efficient*. Otherwise it is unefficient.

3 Results

In this section there are efficiencies evaluated by the both models for all regions for years 2010-2015. Table 4 contains results of output-oriented model BCC-O. There are only values of $\eta^\star$, not vectors $\vec{\lambda}^\star, \vec{s}_+^\star, \vec{s}_-^\star$ found in second phase of optimization. In the same way there are in Table 5 presented results of input-oriented model BBC-I. On the last rows of the both tables there is a number of efficient regions in particular year.

BCC-O Model

Overall the regions have very high efficiency. During the whole period 4 of 14 regions were efficient (A, J, K, P). Three regions were mostly efficient (E, H, L) and the rest was unefficient.

$\eta^\star$	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
A	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
B	0,9930	1,0000	0,9941	0,9925	0,9847	0,9869
C	0,9922	0,9932	0,9929	0,9900	0,9874	0,9867
E	1,0000	1,0000	0,9932	1,0000	0,9971	1,0000
H	0,9980	1,0000	1,0000	1,0000	1,0000	0,9980
J	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
K	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
L	1,0000	1,0000	0,9995	0,9947	1,0000	1,0000
M	0,9848	0,9909	0,9860	0,9831	0,9858	0,9866
P	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
S	0,9848	0,9889	0,9872	0,9851	0,9853	0,9883
T	0,9691	0,9787	0,9660	0,9676	0,9692	0,9703
U	1,0000	0,9723	0,9666	0,9674	0,9688	0,9897
Z	0,9863	0,9989	0,9899	0,9898	0,9852	0,9872
Eff. reg.	7	8	5	6	6	6

Table 4 Efficiency BCC-O

BCC-I Model

Using BCC-I model efficiencies have much more higher variability. Only four regions were efficient all the time (A, J, K, P). Mostly are efficient again two regions (H, L), the others are unefficient.

4 Conclusion

To sum up achieved results we can say that:

- BCC-O model produced results with a very low variability. Its efficiencies have very small range (difference between the maximum and minimum), the minimum is always greater than 0,96. So no big change of outputs is necessary to push unefficient units to the efficient frontier. In Figure 1 there are time series of efficiencies given by BBC-O model. Everywhere there is a constant trend, except of region C (Jihočeský), where the efficiency is decreasing.
- BCC-I model provided much more heterogeneous results. The range is greater than 0,61. The only one input was a number of medical devices in the researched regions. Concerning the trends in BCC-I model they are identical like for previous model - see Figure 2. Only region C has a decreasing trend, the others are constant.
- The number of efficient regions is nearly identical for both models in the researched years (see last rows in Table 4 and Table 5) and it can be considered as a constant equal to 7, i. e. 50 % of regions (7 of 14) is efficient.

This paper is a first part of a big research of temporary absence from work which started this year at our department. Our results confirmed conclusions that have been known from former findings of the Institute of Health Information and Statistics of the Czech Republic. However, the work is going on with further inputs and

$\theta^{\star}$	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
A	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
B	0,5780	1,0000	0,5809	0,5860	0,5613	0,5825
C	0,7507	0,7305	0,7393	0,7076	0,6245	0,6498
E	1,0000	1,0000	0,8191	1,0000	0,8887	1,0000
H	0,8705	1,0000	1,0000	1,0000	0,8925	0,9279
J	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
K	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
L	1,0000	1,0000	0,9959	0,9485	1,0000	1,0000
M	0,5116	0,5482	0,5387	0,5259	0,5651	0,5173
P	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
S	0,5182	0,5468	0,5540	0,5262	0,4942	0,5136
T	0,3794	0,3827	0,3844	0,3515	0,3663	0,3649
U	0,6687	0,5948	0,6074	0,5980	0,5998	0,5952
Z	0,6596	0,7308	0,6384	0,6513	0,6919	0,7088
Eff. reg.	6	8	5	6	5	6

Table 5 Efficiency BCC-I

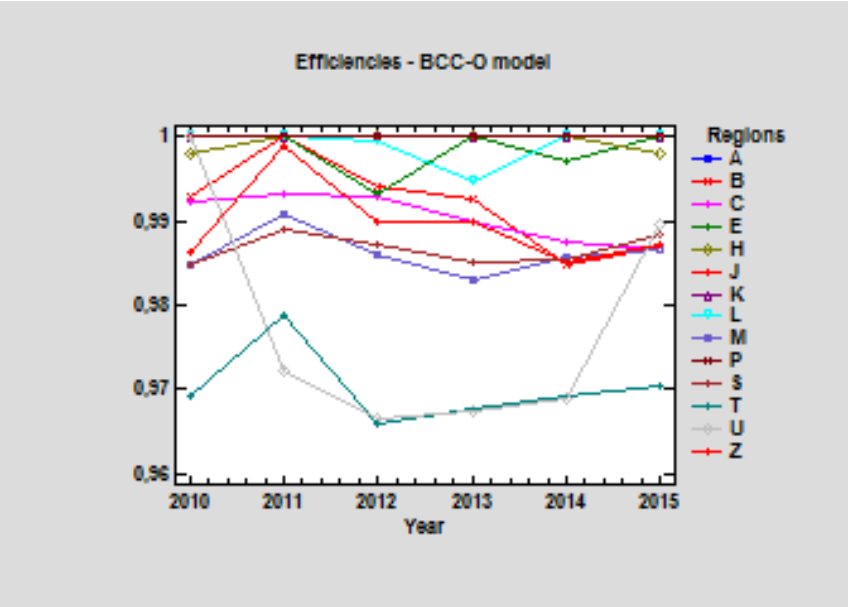


Figure 1 Efficiencies - BCC-O model

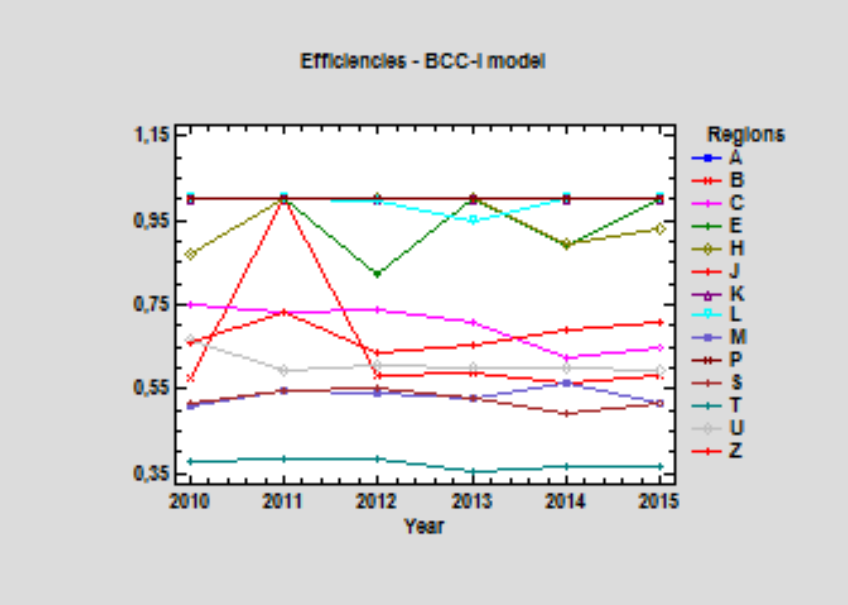


Figure 2 Efficiencies - BCC-I model

outputs and models of DEA.

Acknowledgements

Supported by the grant SGS No. 21187/2017 of Technical University of Liberec.

References

[1] Anděl, J.: *Statistické metody*. MATFYZPRESS, Praha, 2007.

[2] Armingeon, K., and Bonoli, G.: *The Politics of Post-Industrial Welfare States*. Routledge, London, 2006.

[3] Cooper, W. W., Seiford, L. M., and Tone K.: *Data Envelopment Analysis*. Springer, New York, 2007.

[4] Esping-Andersen, G.: *Social Foundations of Postindustrial Economies*. Oxford University Press, Oxford, 1999.

[5] Dlouhý, M., and Jablonský, J.: *Modely hodnocení efektivnosti produkčních jednotek*. Professional Publishing, Praha, 2004.

[6] Jenkins, L., and Anderson, M.: A multivariate statistical approach to reducing number of variables in data envelopment analysis. *European Journal of Operational Research*. **147** (2003), 51-61.

[7] Memariani, A., Farzipoor Saen, R., and Hosseinzadhh, F.: The effect of correlation coefficient among multiple input vectors on the efficiency mean in data envelopment analysis. *European Journal of Operational Research*. **162** (2005), 503-521.

Generalized form of harmonic mean in choosing the optimal value of smoothing parameter in kernel density estimation

Aleksandra Baszczyńska[1]

Abstract. The harmonic mean is one of the Pythagorean means and in particular can be expressed as the reciprocal of the arthimetic mean of the reciprocals. Its generalized form taking into account harmonic mean parameter is used in mathematics, in physics, in electronics, in hydrology, in population genetics and in finance.
The genaralized form of harmonic mean can be used also in choosing the optimal value of smoothing parameter in kernel density function, where two kernel method parameters should be fixed: the kernel function and the smoothing parameter. The shape of kernel function and the value of smoothing parameter influence the final form of the density estimator.
In paper the generalized form of harmonic mean applied in smoothing parameter choice is analysed. Additionally the new method of choosing the optimal value of smoothing parameter is proposed and the properties of the kernel density estimation are regarded with respect to the values of smoothing parameter basing on the generalized form of harmonic means with differeent values of harmonic mean parameter.

Keywords: kernel density estimation, harmonic mean, smoothing parameter.

JEL Classification: C12, C13, C15
AMS Classification: 62G05, 62G10

1 Introduction

Density function or the density estimator is one of the mostly used methods to characterize random variable, univariate as well as multivariate one. It makes possible to present the structure of the data sets. Kernel density estimator is the nonparametric procedure and it is defined for univariate case in the following way:

$$\hat{f}(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right), \tag{1}$$

where: n is the sample size, $x_1,x_2,\ldots,x_n$ is the realization of $X_1,X_2,\ldots,X_n$, K is kernel function and h is the smoothing parameter. Kernel density estimator is widely used but its properties are still investigated ([11], [12]). It may be caused by the diversity of methods for choosing the kernel method parameters such as kernel function K and smoothing parameter h. In the literature the problem of kernel function choice is treated as problem of less importance comparing to the problem of smoothing parameter choice. So, in many practical implementations Gaussian kernel function (density function of standardized normal distribution) is applied. Smoothing parameter can be chosen in a subjective way, basing on reference rules where some assumptions about the population distribution are made or using more sophisticated methods.

The main goal of the paper is to indicate the difficulties connected directly with the choice of smoothing parameter in kernel density estimator. None of the methods for choosing the smoothing parameter regarded in literature cannot be considered as the best one. So, some modification basing on generalized harmonic mean of existing procedures is proposed. Taking into account the goal of the paper the structure is the following: presenting the generalized form of harmonic mean and applying this mean in choosing the values of smoothing parameter in kernel density estimator. Next, the properties of the new method is analyzed by the results of the study indicating such research situations where the appropriate value of parameter of generalized form of harmonic mean should be used.

[1] University of Lodz/Department of Statistical Methods, 90-214 Lodz, 41/43 Rewolucji 1905 St. Poland, albasz@uni.lodz.pl.

2 Generalized form of harmonic mean

The harmonic mean known also as the subcontrary mean belongs to the three classical Pythagorean means group. The means belonging to this group: the arithmetic mean, geometric mean and harmonic mean are characterized by the properties of preservation value, first order homogeneity, invariance under exchange and averaging ([2], [5], [8]). Two of means from this group (harmonic mean and arithmetic mean) can be treated as reciprocal duals of each other for positive arguments while the geometric mean is its own reciprocal dual.

The most known application of harmonic mean is closely connected with the properties of the arguments – it is used when the arguments are rates or rations. But it is not the only application of the harmonic mean. It is used in mathematics in the differential subordination in linear function ([3]). In statistics it can be used in constructing the measure of variability ([1], [9], [10])). In physics it can be used: in averaging of the vehicle speed, in the predicted density of an alloy given the densities of its constituent elements and their mass fractions and in analyzing the resistance of some electrical resistors in parallel and in computing the mean half-life ([7]). In finance the harmonic mean is used for valuing a company in measuring its current share price relative to its per-share earnings (price-earnings ratio)([6]). In geometry it is used when the properties of some geometric figures are is analyzed (for example in triangle the radius of the incircle is one-third of the harmonic mean of the altitudes and in an ellipse the semi-latus rectum is the harmonic mean od the maximum and minimum distances of the ellipse from a focus). In computer science the harmonic mean is used in the evaluation of algorithms and systems in the score of the precision measured by true positives per predicted positive and the recall measured by true positives per real positive. In ecology harmonic mean can be used in measuring the areal distribution of animal activities ([4]).

The unweighted harmonic mean is defined in the following way:

$$H = \left(\frac{\sum_{i=1}^{n} x_i^{-1}}{n} \right)^{-1} = \frac{n \prod_{i=1}^{n} x_i}{\sum_{i=1}^{n} \left(\frac{1}{x_i} \prod_{i=1}^{n} x_i \right)}, \tag{2}$$

where n is the number of arguments and $x_1, x_2, \ldots, x_n$ are real numbers.

The weighted harmonic mean is defined by:

$$H_w = \left(\frac{\sum_{i=1}^{k} w_i x_i^{-1}}{\sum_{i=1}^{k} w_i} \right)^{-1} = \frac{\sum_{i=1}^{k} w_i \prod_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} \left(\frac{1}{w_i x_i} \prod_{i=1}^{k} w_i x_i \right)}, \tag{3}$$

where $w_1, w_2, \ldots, w_k$ is a set of weights.

For the special case of two positive numbers x_1 and x_2, when $x_1 + x_2 \neq 0$, the unweighted harmonic mean is the following:

$$H = \frac{2 x_1 x_2}{x_1 + x_2}. \tag{4}$$

The generalized form of the harmonic mean H_G of two positive numbers x_1 and x_2 is given as (com. [3]):

$$H_G = \frac{x_1 x_2}{x_2 + \beta (x_1 - x_2)} \quad , \tag{5}$$

where: $x_1 + x_2 \neq 0$, $\beta \in (0, 1)$. The parameter β is the harmonic mean parameter that controls the weights assign to x_1 and x_2. For $\beta = \frac{1}{2}$ the generalized form of harmonic mean results in formula (2).

3 Smoothing parameter in kernel density estimation

The calculation of the kernel density estimator requires the value of smoothing parameter. Quick and dirty methods of smoothing parameter choice play the special role – they are simple to calculate and easy to use and moreover, these methods are very often used as the starting point in more sophisticated procedures.

In reference rule the smoothing parameter is such a value for which the asymptotic mean integrated squared error between the kernel density estimator and the density function takes the minimum value, assuming that unknown density function is the normal one with the scale as that estimated for the underlying density. As mentioned in the literature this procedure gives the proper results when the data are normally distributed or close to it. Reference rule smoothing parameter is the following:

$$h_{RR} = 0.9An^{-\frac{1}{5}}, \tag{6}$$

where $A = \min\left(\hat{\sigma}, \frac{\hat{R}}{1.34}\right)$, $\hat{\sigma}$ and $\hat{R}$ are the sample standard deviation and sample semi-interquartile range, respectively.

In maximal smoothing procedure, the smoothing parameter is such a value for which the asymptotic mean integrated squared error between the kernel density estimator and the density function takes the minimum value, assuming that unknown density function is the beta distribution $B(4,4)$. It can be shown that for density with the standard deviation σ the beta distribution is the upper bound for the estimated densities. Maximal smoothing parameter is the following:

$$h_{MS} = \frac{3}{35^5}\hat{\sigma}\left(\frac{R(K)}{\kappa_2^2}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}, \tag{7}$$

where $R(K) = \int_{-\infty}^{+\infty} [K(x)]^2 dx$ and $\kappa_2 = \int_{-\infty}^{+\infty} u^2 K(u) du$.

Details of the methodology of choosing smoothing parameter methods can be found in literature ([11], [12]).

4 Results of the study

To investigate the properties of the new method of choosing the smoothing parameter in kernel density estimation the study was conducted. Basing on the samples, the kernel density estimators are constructed taking into account appropriate kernel function and values of smoothing parameters calculated as the harmonic mean of generalized form with different values of harmonic mean parameter. The values of smoothing parameters are compared to analyze the properties of generalized form of harmonic mean with different values of harmonic mean parameters. In this way it is possible to indicate the best values of harmonic mean parameter in a special scientific research.

In the study the following data sets are regarded:

1. data set A: samples drawn from the population with density function $f(x) = wf_1(x) + (1-w)f_2(x)$, where $w = 1$ and $f_1(x)$ is density function $N(0,1)$, sample size n =10, 20, 30, 40, 50, 60, 70, 80, 90 and 100;
2. data set B: samples drawn from the population with density function $f(x) = wf_1(x) + (1-w)f_2(x)$, where w = 0.5 and $f_1(x)$ is density function $N(0,1)$, $f_2(x)$ is density function $N(10,1)$, sample size n =10, 20, 30, 40, 50, 60, 70, 80, 90 and 100;
3. data set C: samples drawn from the population with density function $f(x) = wf_1(x) + (1-w)f_2(x)$, where w = 0.8 and $f_1(x)$ is density function $N(0,1)$, $f_2(x)$ is density function $N(10,1)$, sample size n =10, 20, 30, 40, 50, 60, 70, 80, 90 and 100;
4. data set D: values of sales and net profit in 2015 of Polish enterprises with incomes of 250 million PZ, n =166.

The choice of distributions of data sets A, B and C is motivated by the attempt to take into regard unimodal and symmetric populations (data set A) as well as bimodal (data sets B and C) and asymmetric (data set C) ones. Different sample sizes widen the study to the situations where small, medium and large samples are considered.

For every sample the kernel density estimator is calculated with Gaussian kernel function and smoothing parameter using the method of reference rule and maximal smoothing. Next, the values of generalized harmonic mean method for smoothing parameter are calculated for the harmonic mean parameter β =0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 with $x_1 = h_{RR}$ and $x_2 = h_{MS}$. Tables 1-3 present smoothing parameter values for chosen sample sizes and different values of harmonic mean parameter.

n	h_{RR} h_{MS}	$h_{GH\beta}$								
		$\beta = 0.9$	$\beta = 0.8$	$\beta = 0.7$	$\beta = 0.6$	$\beta = 0.5$	$\beta = 0.4$	$\beta = 0.3$	$\beta = 0.2$	$\beta = 0.1$
10	0.3210 0.3466	0.3234	0.3258	0.3288	0.3308	0.3333	0.3359	0.3385	0.3412	0.3439
30	0.4621 0.4990	0.4655	0.4690	0.4726	0.4762	0.4798	0.4836	0.4873	0.4912	0.4951
50	0.4770 0.5152	0.4806	0.4842	0.4879	0.4916	0.4954	0.4992	0.5031	0.5071	0.5111
70	0.3935 0.4249	0.3965	0.3994	0.4024	0.4055	0.4086	0.4118	0.4150	0.4182	0.4216
90	0.3388 0.3659	0.3414	0.3439	0.3465	0.3492	0.3519	0.3546	0.3573	0.3602	0.3630
100	0.3974 0.4292	0.4003	0.4034	0.4064	0.4095	0.4127	0.4159	0.4191	0.4224	0.4257

Table 1 Values of smoothing parameter for data set A

n	h_{RR} h_{MS}	$h_{GH\beta}$								
		$\beta = 0.9$	$\beta = 0.8$	$\beta = 0.7$	$\beta = 0.6$	$\beta = 0.5$	$\beta = 0.4$	$\beta = 0.3$	$\beta = 0.2$	$\beta = 0.1$
10	3.2008 3.4567	3.2247	3.2489	3.2735	3.2985	3.3238	3.3496	3.3757	3.4023	3.4293
30	2.6976 2.9132	2.7177	2.7381	2.7589	2.7799	2.8013	2.8230	2.8450	2.8674	2.8901
50	2.4084 2.6009	2.4264	2.4446	2.4631	2.4819	2.5010	2.5203	2.5400	2.5600	2.5803
70	2.4222 2.6159	2.4403	2.4586	2.4772	2.4961	2.5153	2.5348	2.5546	2.5747	2.5951
90	2.1972 2.3728	2.2136	2.2302	2.2471	2.2642	2.2816	2.2993	2.3172	2.3355	2.3540
100	2.1581 2.3306	2.1742	2.1905	2.2071	2.2239	2.2410	2.2584	2.2760	2.2939	2.3121

Table 2 Values of smoothing parameter for data sets B

n	h_{RR} h_{MS}	$h_{GH\beta}$								
		$\beta = 0.9$	$\beta = 0.8$	$\beta = 0.7$	$\beta = 0.6$	$\beta = 0.5$	$\beta = 0.4$	$\beta = 0.3$	$\beta = 0.2$	$\beta = 0.1$

10	0.6096 0.6584	0.6142	0.6188	0.6235	0.6282	0.6330	0.6380	0.6429	0.6480	0.6531
30	0.5888 0.6360	0.5932	0.5977	0.6022	0.6068	0.6115	0.6162	0.6210	0.6259	0.6309
50	0.8166 0.8819	0.8227	0.8288	0.8351	0.8415	0.8480	0.8545	0.8612	0.8680	0.8749
70	0.7493 0.8092	0.7549	0.7606	0.7663	0.7722	0.7781	0.7841	0.7903	0.7965	0.8028
90	0.3388 0.3659	0.3414	0.3439	0.3465	0.3492	0.3519	0.3546	0.3573	0.3602	0.3630
100	0.6494 0.7013	0.6542	0.6591	0.6641	0.6692	0.6743	0.6796	0.6849	0.6903	0.6957

Table 3 Values of smoothing parameter for data sets C

Figure 1 presents kernel density estimators for sales (on the left) and net profit (on the right) in 2015 of Polish enterprises with incomes of 250 million PZ. Basing on our previous experience of the right asymmetry of regarded variables, the harmonic mean parameter is taken $\beta = 0.1$ and Gaussian kernel function is implemented.

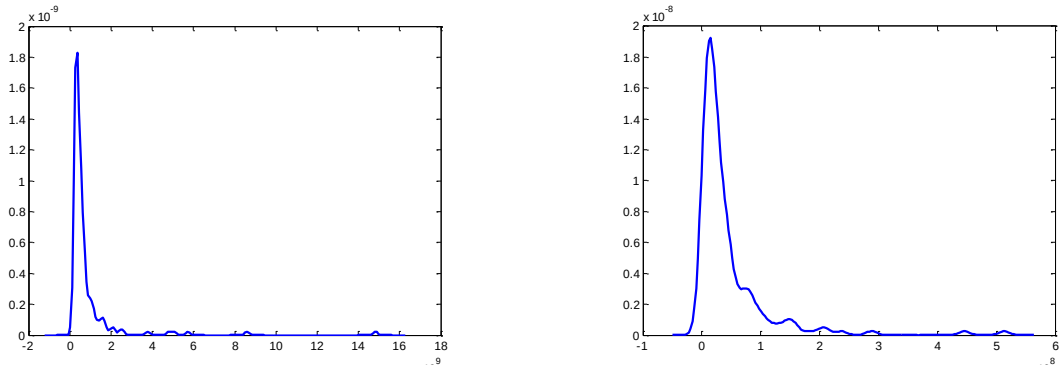


Figure 1 Kernel density estimators for the data set D

5 Conclusions

When samples are drawn from symmetric and unimodal population the differences between values of smoothing parameters calculated with reference rule and maximal smoothing are significantly smaller than in case of asymmetric or bimodal populations. It can indicate that two methods of choosing the smoothing parameter in kernel density estimation regarded in the study are of different nature. The smoothing parameters from reference rule and maximal smoothing are similar only when population is normal. Moreover, for samples from bimodal populations these differences are bigger when population is symmetric. It indicates that reference rule and maximal smoothing cannot be treated as universal methods of choosing the smoothing parameter. Hence, there is a necessity for calculating such a value of smoothing parameter that would unite regarded methods. Smoothing parameter constructed as generalized harmonic mean with the possibility of assigning different weights to reference rule or maximal smoothing parameter can play such a role. It should also be noticed that in general the bigger the sample size the smaller the differences between two regarded methods.

In the constructing of smoothing parameter by generalized harmonic mean the following issues should be considered:

1. for samples from symmetric unimodal population (or from populations close to symmetric) the harmonic mean parameter should be rather close to one, then the reference rule would have bigger importance;
2. for samples from bimodal populations the bigger harmonic mean parameter indicates the bigger differences between values of smoothing parameters, then the medium value of harmonic mean parameter is preferred;
3. for samples from asymmetric bimodal populations the harmonic mean parameter should be close to zero, then the method of maximal smoothing is of greater importance.

Taking into consideration the results of the study as well as the additional information of the character of the variable regarded in data set D (for example the information about right-sided asymmetry from previous researches) in constructing the kernel density estimator for the sales and net profit of Polish enterprises with incomes of 250 million PZ in 2015 the harmonic mean parameter is 0.1. Kernel density estimators (with the appropriate values of smoothing parameter) presented in Figure 1 shows the most important features of regarded random variables. For example, the strong asymmetry of variables disables usage of the classical statistical methods in next, more sophisticated statistical analysis.

Further analysis should be devoted to the extension of regarded method for example in calculating the so-called the acceptable interval of smoothing parameter. Because of the simplicity and universalism of the method of generalized harmonic mean it can be used also for other more sophisticated method of choosing the smoothing parameter in kernel density estimation. Additionally, the other kernel functions can be used in kernel estimator.

References

[1] Brown, M.: A measure of variability based on the harmonic mean, and its use in approximations. *The Annals of Statistics* **13,** 3 (2016), 1239-1243.

[2] Burić, T.: Asymptotic behaviour of the iterative Pythagorean means. *Rad Hazu. Matematičke znanosti* **19**, 523 (2015), 117-127.

[3] Chojnacka, O., Lecko, A.: On the differential subordination of harmonic mean to a linear function. *Bulletin de la Société et des Lettres de Łódź, série: Recherches sur les Déformations* **LXVI,** 2 (2016), 118-127.

[4] Dixon, K., Chapman, J.: Harmonic mean measure of animal activity areas. *Ecology* **61(5)** (1980), 1040-1044.

[5] Furuichi, S.: Operator inequalities among arithmetic mean, geometric mean and harmonic mean. *Journal of Mathematical Inequalities* **8,** 3 (2014), 669-672.

[6] Kijima, M.: The generalized harmonic mean and a porfolio problem with dependent assets. *Theory and Decision* **43** (1997), 71-87.

[7] Lam, F., Hung, C., Perrier, D.: Estimation of variance for harmonic mean half-lives. *Journal of Pharmaceutical Sciences* **74**, 2(1985), 229-231.

[8] Meyer, B.: Some inequalities for elementary mean values. *Mathematics of Computation* **42,** 165 (1984), 193-194.

[9] Padois, T., Doutres, O., Sgard, F., Berry, A.: On the use of geometric and harmonic means with the generalized cross-correlation in the time domain to improve noise sourse maps. *Journal of the Acoustical Society of America* **140**, EL56 (2016), http://dx.doi.org/10.1121/1.4955007 [15.04.2017].

[10] Sharma, R.: Some more inequalities for arthimetic mean, harmonic mean and variance. *Journal of Mathematical Inequalities* **2,** 1 (2008), 109-114.

[11] Silverman, B.: *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

[12] Wand, M., Jones, C.: *Kernel Smoothing*, Chapman and Hall, London, 1995.

A Medium-scale DSGE Model with Labour Market Frictions

Jakub Bechný[1], Osvald Vašíček[2]

Abstract. The aim of our research is to find the most appropriate approach to the labour market within a DSGE model estimated on the Czech Republic's data. We use slightly modified medium-scale small open economy model originally proposed by Sheen and Wang [10]. This model incorporates a standard set of rigidities such as investment adjustment cost, habit formation, external risk premium, and incomplete exchange rate pass-through. The model allows for two approaches to the labour market - the real wage rigidity specified in spirit of the search and matching models, and the nominal wage rigidity a la Calvo wage setting. The model is estimated on a set of fourteen observable variables, including the unemployment rate. Our research is in the early stage, however, preliminary results suggest that the real wage rigidity is more appropriate modelling approach. Empirical performance of our model is illustrated by means of the variance decomposition and historical shock decomposition of selected observable variables.

Keywords: DSGE model, small open economy, labour market frictions, unemployment rate.

JEL classification: C32, E17
AMS classification: 91B51

1 Introduction

Standard dynamic stochastic general equilibrium (DSGE) models does not explicitly incorporate unemployment. Labour market activity is in majority of current DSGE models captured either by varying hours worked, or by choice whether or not to participate on the labour market at all. Authors such as Blanchard [3] criticize this approach, because unemployment rate is an important indicator of the aggregate economic activity, and also brings a negative social consequences. It may be interesting to investigate which shocks account for fluctuations in unemployment, how shocks from labour market propagate throughout economy, and how is labour market affected by monetary policy. Our research is therefore focused on incorporation of labour market with explicitly modelled unemployment rate into a standard medium-scale DSGE model.

As the first step in our research, which is summarized in this short paper, we decided to replicate a medium scale DSGE model of Sheen and Wang [10]. This model allows for two alternative approaches to the labour market - the real wage rigidity specified in spirit of the search and matching models, and the nominal wage rigidity a la Calvo [5] and Yun [11] sticky wage setting. The model of Sheen and Wang [10] was originally estimated on Australian data. We introduce three small modifications of the model (related to the open economy features and monetary policy), and estimate it using data for the Czech Republic from 2004Q3 to 2017Q1. The estimated model is then used for evaluation whether nominal or real wage rigidity is supported by the Czech data, and for quantification of hiring costs in the Czech Republic. To illustrate the overall ability of the model to explain the data we also decompose variances of key observable variables, and we provide historical shock decomposition of the Czech Republic's unemployment rate. We also present estimates of the productivity growth, which is the only source of the long run economic growth in our model. The remainder of this short paper is organized as follows. Section 2 sketches our model with focus on the labour market block. Selected results of our empirical analysis are then presented in section 3.

2 Model

Structure of the Sheen and Wang [10] model is standard. It combines the closed economy setting of Christiano *et al.* [7], the small open economy features of Adolfson *et al.* [1], and it adds more elaborate labour market block with explicitly modelled unemployment. All real variables of the model share a common trend due to a unit-root technology shock, and the model exhibits a balanced growth path steady state. There are four types of representative firms in the model economy: domestic goods-producing firms, consumption importers, investment

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, jakub.bechny@gmail.com
[2]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, osvald.vasicek@econ.muni.cz

importers, and exporters. Following Calvo [5] and Yun [11] the model assumes that prices of all firms are sticky, and each firm sets price based on its markup and real marginal cost. Sticky prices of import and export firms also allow for incomplete exchange rate pass-through. Capital services and labour are used only by domestic goods-producers, depending on the real wage and the rental rate.

Households attain utility from leisure, real balances, and from consumption of both domestic and imported consumption goods, subject to habit formation. Households can hold both domestic and foreign bonds, and their investment decisions yield the uncovered interest rate parity condition with risk premium. Households also own the physical capital stock, and they can vary capital services in three ways. They can accumulate the physical capital and pay investment adjustment cost, they can vary the utilization rate of physical capital at cost, or they can trade installed capital among themselves. Behaviour of a central bank is characterized by an interest rate rule, which responds to the CPI inflation rate, the output gap, and the real exchange rate.

In order to more precisely depict structure of the Czech economy, we introduce several modifications of the original Sheen and Wang [10] model. Firstly, we use a modified risk-adjusted uncovered interest parity (UIP) condition of Adolfson *et al.* [2], which allows for higher persistence and volatility of the real exchange rate. Secondly, instead of the original reduced form VAR(1) specification of the foreign economy block, we follow Pedersen and Ravn [9] and use more structural specification. Our foreign economy consists of monetary policy rule, New Keynesian Phillips curve, and IS curve (analogical to the domestic consumption Euler equation). This specification and the modified UIP condition allow for better identification of open economy structural shocks, which account for a substantial part of observed fluctuations in the Czech economy as we show later. Thirdly, in order to depict changes in inflation target of the Czech National Bank during the investigated time period we follow Adolfson *et al.* [1] and we introduce time varying inflation target into the model. The following two sections of this paper sketch the labour market block of the model, but for detailed description of the optimization problems see paper of Sheen and Wang [10].

2.1 Labour market flows

The labour force is in the model normalized to 1. Our model assumes that a fraction δ of the employees N_t is each period separated from existing job. Denoting H_t aggregate hiring, the evolution of hiring is given by

$$H_t = N_t - (1-\delta)N_{t-1} \tag{1}$$

Domestic goods-producers pay a real cost of hiring a new worker g_t, which is proportional to a current labour market tightness x_t

$$g_t = \epsilon_t B x_t^{\vartheta} \zeta_t^x \tag{2}$$

where ϵ_t is a temporary AR(1) technology shock, B is parameter that measures the scale of the hiring cost, ϑ is elasticity of hiring cost to labour market condition, and ζ_t^x is AR(1) hiring cost shock. Blanchard and Gali [4] show that this form of the hiring cost can be derived from a matching function which is homogeneous of degree one. The labour market tightness x_t is defined as the ratio of the hires to the number of unemployed before hiring proceeds at time t, where the unemployment is given by $U_{t-1} = 1 - N_{t-1}$

$$x_t = \frac{H_t}{U_{t-1} + \delta N_{t-1}} \tag{3}$$

2.2 Real and nominal wage rigidity

Our model allows for real and nominal wage rigidities as alternatives. The real wage rigidity is in spirit of Hall [6] constructed as the weighted average of the lagged real wage w_{t-1} and the equilibrium Nash-bargaining wage w_t^*

$$w_t = f w_{t-1} + (1-f) w_t^* \tag{4}$$

and f represents the degree of the rigidity. The equilibrium wage resulting from the Nash-bargaining between firms and workers is in turn given by

$$w_t^* = g_t + \frac{\zeta_t^N N_t^{\sigma_L}}{\psi_t^z} - (1-\delta)\beta E_t \left[\frac{\psi_{t+1}^z}{\psi_t^z \mu_{t+1}^z \pi_{t+1}^d} (1 - x_{t+1}) g_{t+1} \right] \tag{5}$$

where ζ_t^N is a stationary AR(1) labour supply preference shock, ψ_t^z is marginal utility from real income, μ_t^z is AR(1) process describing the growth rate of the permanent technology process, π_t^d is the gross domestic inflation rate, σ_L represents the inverse Frisch elasticity of labour supply, and β is households' discount factor.

Alternatively, nominal wage rigidity is constructed by assuming that labour is differentiated and organized (possibly in unions), and follows the Calvo - Yun type of wage setting. Suppose that the household j is allowed to

change its real wage at $w_{j,t}^{new}$ in period t, and that it is not allowed to change its wage during s periods ahead. The utility maximization problem of this household implies the following first order condition

$$E_t \sum_{s=0}^{\infty} (\beta\xi_w)^s N_{j,t+s} \left[-\zeta_{t+s}^N N_{j,t+s}^{\sigma_L} + w_{j,t}^{new} \frac{\psi_{t+s}^z}{\lambda_{t+s}^w} \frac{P_{t+s-1}^c}{P_{t-1}^c} \frac{P_t^d}{P_{t+s}^d} \right] = 0 \tag{6}$$

where ξ_w is the Calvo - Yun probability that the wage will not be reoptimized s periods ahead, λ_t^w is wage markup which follows AR(1) process, P_t^c is the aggregate price for consumption goods, and P_t^d is aggregate domestic price level. Equation (6) can be log linearised, rewritten recursively, and combined with an aggregate wage index that connects $w_{j,t}^{new}$ with current and lagged real wage. This algebra would give us the Hybrid New Keynesian Wage Phillips Curve.

3 Empirical analysis

For purposes of the empirical analysis we log-linearise all 64 equations of our model, solve the whole system numerically using Dynare, and estimate its parameters by using the Bayesian approach. Our model is estimated using quarterly time series for the Czech Republic from 2004Q3 to 2017Q1. We use data from the Czech National Bank's ARAD database, including CPI inflation, nominal cash rate (PRIBOR), real GDP, consumption, investment, imports and exports, unemployment rate, real wages, inflation target, and CZK/EUR nominal exchange rate. Foreign economy sector is proxied by the Euro Area nominal cash rate, real GDP, and by CPI inflation. All real variables in our theoretical model share the same stochastic trend due to growth in technology. We therefore directly identify the growth component by mapping our model variables to the first differences of the logged observable variables.

We do not report estimates of the parameters here to save space. With the estimated model in hand we perform several empirical exercises. Firstly, we quantify whether the hiring costs are significant in the data and evaluate whether nominal or real wage rigidity is supported by the data. Secondly, we compute the forecast variance decomposition of the various shocks on key observable variables. Thirdly, we analyse the recent development of the Czech Republic's economy by means of the historical shock decomposition. And finally, we provide estimates of the permanent technology process, which is in our model the only source of the long term economic growth.

3.1 Wage rigidities and hiring costs

	RWR, HC	RWR, NHC	NWR, HC	NWR, NHC
Log-Likelihood	-1537.80	-1548.91	-1572.82	-1575.12
BF(Null hypothesis: NHC)	11.11	-	2.3	-
BF(Null hypothesis: NWR)	35.02	26.24	-	-
RWR: Real wage rigidity		NWR: Nominal wage rigidity		
HC: Hiring cost		NHC: No hiring cost		BF: Bayes factor

Table 1 Comparison of Models

We estimated four alternative variants of our model: the model with and without hiring costs, and with either real or nominal wage rigidity. Following Kass and Raftery [8], we use the Bayes factor as the model selection criterion.[3] Results of the model selection are summarized in Table 1. Comparison of the posterior log-likelihoods shows that the model with hiring costs and real wage rigidity fits the data best (with value of log likelihood -1537.8). The Bayes factors show that the data provide very strong evidence against the model with the nominal wage rigidity, regardless of the existence of hiring costs (with Bayes factor of 35.02, respectively 26.24). In case of the real wage rigidity specification, data also provide very strong evidence against the model without hiring costs (BF 11.11). In case of the nominal wage rigidity specification, evidence against the model without hiring costs is only positive (BF 2.3).

One of key parameters in our model is B, which determines steady state level of hiring costs, see Equation (2). The estimated posterior mean of this parameter is 0.19, with 90% credible interval between 0.13 and 0.26. This implies that in the balanced growth path steady state the hiring costs account for 0.1 % of the real GDP, with 90% credible interval between 0.07 % and 0.14 % of the GDP. Model with hiring costs is thus strongly supported by

[3]Denote $p(Data \mid H_0)$ as the log-likelihood of data conditional on the null hypothesis H_0, and $p(Data \mid H_1)$ as the log-likelihood conditional on the alternative model H_1. The logarithm of the Bayes factor is given by $B_{10} = p(Data \mid H_1) - p(Data \mid H_0)$. Kass and Raftery [8] argue that a value of B_{10} between 1 and 3 provides a positive evidence against model H_0, a value between 3 and 5 provides strong evidence, and a value greater than 5 provides very strong evidence against H_0.

the data according to the Bayes factor, but relative size of the hiring costs in the economy is rather limited. Note also that these hiring costs are conceptualized as the real costs of firm associated with searching and hiring of the labour, and that they do not account for e.g. payroll taxes.

3.2 Variance decomposition

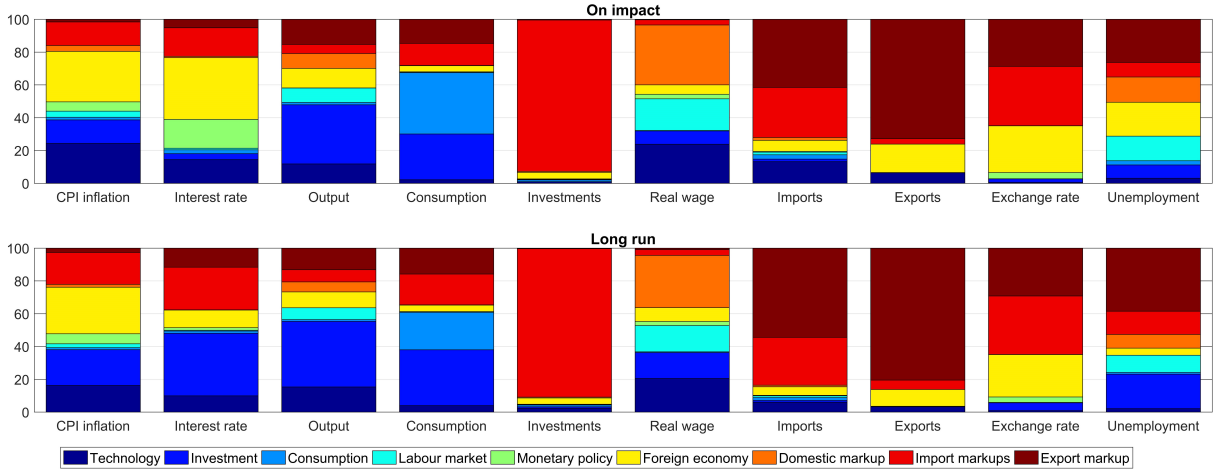


Figure 1 Variance decomposition

Figure 1 shows the forecast variance decomposition on impact and in the long run (20 quarters) of shocks on observable variables from our model.[4] We aggregate the effects of temporary and permanent technology shocks into one single Technology shock, and aggregate all shocks related to the foreign economy[5] to the Foreign economy shock. Imported consumption and imported investment markups are aggregated into single Import markup. The remaining measures represent shocks to the Investments, Consumption preference, Labour market (aggregated labour supply and hiring cost shock), Monetary policy (aggregated shocks to the monetary policy rule and to the inflation target), Domestic markup, and Export markup. Our analysis reveals several interesting results.

Overall, shocks related to open economy features (foreign economy shocks, import and export markups) dominate variances of almost all variables both in the short and long run. This result may reflect fact that the Czech Republic is small open economy whose condition crucially depends on performance of export and import sectors. For example, these shocks account approximately for 60 % of nominal interest rate variance in the short run. In the long run, they account for more than 90 % of the investments, imports, exports, and exchange rate variation. Investment specific shock is another dominant shock, in the long run accounting for more than 30 % of the interest rate output and consumption variation. Monetary policy shocks have only limited impact - they account for 15 % of nominal interest rate's short run variation, and for 5 % of exchange rate's and inflation's movements. This indicates that the Czech National Bank has not caused large surprises to the Czech economy, even in the short run.

Regarding the labour market, almost 55 % of the variance of unemployment is both in the short and long run explained by the open economy shocks. Investment shocks play a significant role especially in the long run, explaining almost 20 % of the variance of unemployment. Labour market shocks then account for 15 % of unemployment's short run variation, and of 10 % variation in the long run. Variability of the real wage is mainly explained by technology shocks (20 %), labour supply shocks (20 %), and domestic markup (40 %) both in short and long run. Labour market shocks also explain almost 10 % of the real GDP in both short and long run.

3.3 Historical shock decomposition

In this section we discuss the main driving forces of unemployment identified by our model by means of the historical shock decomposition visualised in Figure 2. Thick line in this figure depicts demeaned unemployment rate. For purposes of this exercise we aggregate the effects of all markup shocks into single Markup shock, and effects of the consumption preference and monetary policy shocks are aggregated into single Demand shock. Technology and Investment shocks are aggregated into single technology shock. Foreign economy and Labour market shocks are defined as in the previous section.

Regarding the 2008-2009 crisis, the unemployment rate almost continuously incriesed from 4.3 % in 2008Q3

[4] As is discussed by Adolfson *et al.* [1], because the unit root technology shock is the only permanent shock in our model, it would account for all fluctuations in the limit. The long run is therefore selected as 20 quarters.

[5] Foreign monetary policy, inflation and demand shocks, foreign technology, and risk premium shock.

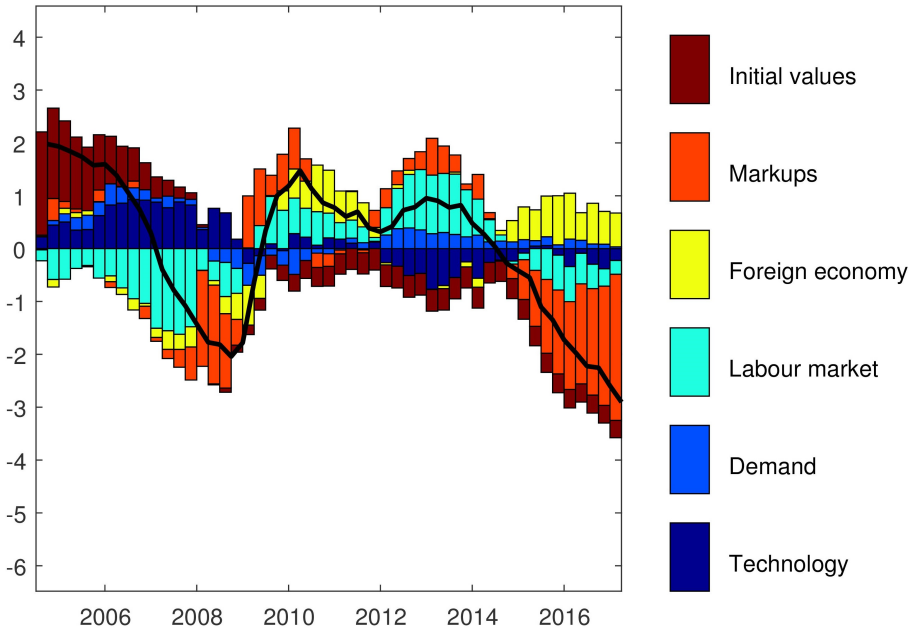


Figure 2 Historical decomposition of unemployment rate (demeaned)

to 7.8 % in 2010Q1. Our model attributes the initial increase of the unemployment to the adverse markup shocks (with dominant role of export and import markups), and to the labour market shocks. These shocks were followed by the adverse open economy shocks in the subsequent periods. Note also that the demand shocks were mildly positive during this recession. Our model thus suggests that rise of the unemployment during 2008-2009 crisis can be to large extend attributed to situation in the sectors related to the foreign economy development, which seems as quite plausible explanation.

During the 2012-2013 recession the unemployment rate started at 6.5 % in 2011Q4 and peaked at 7.3 % by 2012Q4. In contrast with the 2008-2009 crisis, our model attributes this development almost solely to the adverse labour market shocks. Note also that demand shocks were negative during this crisis, with dominant role of the consumption preference shock. Adverse labour market and demand shocks may be attributed to the overall negative sentiment in the economy and also to the restrictive fiscal policy of the former government.

The unemployment rate continuously decreased from 7 % by 2013 to its historically lowest levels, 3.4 % by 2017Q1. Our model attributes this development to the markup shocks, with dominant role of the export markup. These export markup shocks depict increased competitiveness of the domestic exporters, and may be attributed to the exchange rate commitment of the Czech National Bank from November 2013. Note also that the foreign economy shocks, which capture development in the foreign economy block itself, had a negative impact on the Czech unemployment rate during the discussed time period.

3.4 Permanent technology growth

All real variables in our model share a common stochastic trend due to a unit root permanent technology (or productivity) shock. Growth rate of this permanent technology, which is in our model the only source of the long run economic growth, can be estimated and is visualised in Figure 3. We present 90% credible intervals for this estimate, and compare the estimated growth rate of permanent technology with the observable quarter-on-quarter real GDP growth. Note that the 2008-2009 recession was associated with significant drop in the technology growth, while the 2012-2013 recession brought only very moderate deterioration of the technology growth. From 2014 onwards, our estimates imply the permanent technology growth around 1 % per quarter.

4 Conclusion

This paper presented results of estimation of a medium scale open economy DSGE model with labour market frictions. The model was estimated on the Czech Republic's data from 2004 to 2017. Our main findings can be summarized as follows. Firstly, we compared four alternative specifications of the labour market block (with real or nominal wage rigidity, and with or without hiring costs). By using the Bayesian approach to model selection we showed that the real wage rigidity with the hiring costs is the most appropriate specification of the Czech labour market. Secondly, we assessed empirical performance of the model by means of the variance decomposition

and historical shock decomposition. We showed that variances of the key macroeconomic variables are strongly dominated by shocks related to open economy features. We showed that increase in unemployment during the 2008-2009 crisis can be attributed mainly to the foreign economy development, while the 2012-2013 recession is associated with adverse labour market and demand shocks. Historically low unemployment of 3.4 % by 2017Q1 then may be attributed to the export markup shocks. And finally, we presented estimates of the permanent technology process. We showed that the 2008-2009 recession was so severe that it affected even the permanent technology, the only source of the long run economic growth in our model.

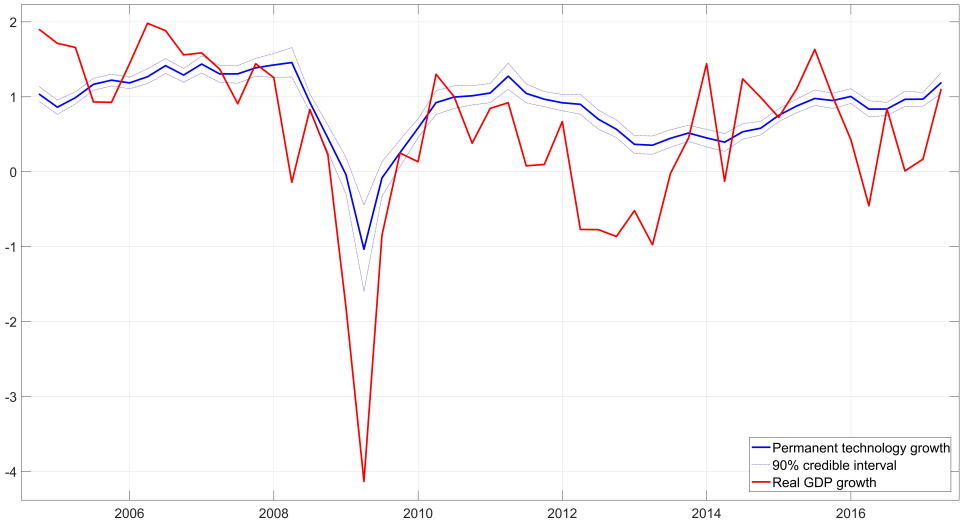


Figure 3 Permanent technology growth in Czech Republic

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0987/2016. This support is greatly appreciated.

References

[1] Adolfson, M., Laséen, S., Lindé, J., and Villani, M.: Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through, *Journal of International Economics* **72.2** (2007), 481–511.

[2] Adolfson, M., Laséen, S., Lindé, J., and Villani, M.: Evaluating an Estimated New Keynesian Small Open Economy Model, *Journal of Economic Dynamics and Control* **32.8** (2008), 2690–2721.

[3] Blanchard, O.: The State of Macro, *Annual Review of Economics* **1.1** (2009), 209–228.

[4] Blanchard, O. and Gali, J.: Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment, *American Economic Journal: Macroeconomics* **2.2** (2010), 1–30.

[5] Calvo, G.: Staggered Prices in a Utility-Maximizing Framework, *Journal of Monetary Economics* **12.3** (1983), 383–398.

[6] Hall, R. E.: Employment Fluctuations with Equilibrium Wage Stickiness, *The American Economic Review* **95.1** (2005), 50–65.

[7] Christiano, L. J., Eichenbaum, M., and Evans, Ch., L.: Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy, *Journal of Political Economy* **113.1** (2005), 1–45.

[8] Kass, R. E. and Raftery, A. E.: Bayes Factors, *Journal of the American Statistical Association* **430** (1995), 773–795.

[9] Pedersen, J. and Ravn, S. H.: *What Drives the Business Cycle in a Small Open Economy? Evidence from an Estimated DSGE Model of the Danish Economy*. Danmarks Nationalbank Working Papers No. 88 (2013).

[10] Sheen, J. and Wang, B. Z.: *An Estimated Small Open Economy Model with Labour Market Frictions*. Dynare Working Papers Series no. 35 (2014).

[11] Yun, T.: Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles, *Journal of monetary Economics* **37.2** (1996), 345–370.

Price comparison sites and their influence on e-commerce processes

Ladislav Beranek[1], Petr Hanzal[2]

Abstract. Customers often use environment price comparison sites (e.g., heureka.cz, or zbozi.cz) when they want to buy some product or service over the Internet. These sites help them while searching of some product, give them an overview on prices, and also affect consumers' perception of the risks associated with online buying. The purpose of our paper is to suggest a model which would explain basic processes of online shopping environment with the tools of network analysis. We look at the relation of shoppers and e-shops as bipartite network. Our paper should answer following questions as: What communities are forming and how? Can we determine conditions on network topology? We use simulation data for this purpose, and compare results with data from other resources.

Keywords: E-commerce; Price comparison sites; Network analysis; Simulation.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Internet has proved to be an important platform for various business transactions. According to the Czech Statistical Office (2016), retail e-commerce has grown in the Czech Republic nearly to US$3.3 billion in 2015 surpassing the sales of US$2.7 billion in year 2014 and it was estimated to increase to US$4.5 billion in 2017. Simultaneously with the development of electronic business, related business applications have been developed, and many various methods very used while doing so [24, 25]. Price comparison sites that allow to compare prices in different e-shops are ones of the most successful (see Fig. 1). These sites are also known as price comparison sites, shopbots, or Internet shopping agents. Online buyers use them to gain information about price or respective shops. They reduce buyers' search costs and help their decision-making by providing price comparison information, which is seldom present in the physical retail shopping context [7].

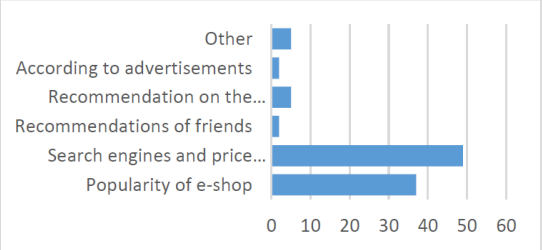


Figure 1 How do you mostly choose some e-shop, when you are going to buy a product or service on Internet? (in per cent, our own research, 2016).

Price comparison sites were studied mostly from the perspective of the impact of the existence of price comparison sites on price of products and services and on sensitivity of online shoppers to price [2]. Degeratu et al. [10] reported that the existence of price comparison sites increased price competition and sensitivity of buyers to price. In this paper however, we want to look at the processes of online shopping with optics of network analysis.

Social and business networks are an increasingly important area of research attention in many disciplines [5, 11, 15]. However, stable equilibriums and models have been mainly focused on, while their dynamics and evolution have received limited research attention. One of the major challenges is to better understand, predict and control their dynamics, including how they form, evolve and shape their behaviors and performances [21, 23]. Significant progress has been made in e-commerce applications so far, and e-commerce plays a very important

[1] Faculty of Economics, University of South Bohemia, Studentska 13, České Budejovice, Czech Republic beranek@ef.jcu .cz.
[2] Faculty of Economics, University of South Bohemia, Studentska 13, České Budejovice, Czech Republic hanzal@ef.jcu .cz.

role in economy. Large numbers of buyers and sellers interact with each other through transactions on websites [3, 4]. These interactions promote the evolution and shape complex structures of e-commerce market. Going deeply insight into research of e-commerce market network has a profound and lasting significance.

Our research is aimed to develop a model to investigate the processes of online shopping with an existence of a price comparison site with the help of tools of network analysis. We look at the relation of shoppers and e-shops as tripartite network. This is our research problem with the main questions: What communities are forming and how? What relations are prevailing? How does a customer decide based on price comparison sites information which of the e-retailer she buys some product from? These are the main questions we addressed in this document. Accordingly, the objectives of this study are two folds:

1. Building a basic theoretical model, by using network analysis methods to model explain how buyers use price comparison sites when they perform their purchases;

2. Conducting an empirical research to examine the validity and explanatory power of this model.

We use simulation data for this purpose, and compare results with data from other resources.

2 Theoretical background

In this section, we present a network analysis framework for e-commerce market. The network is constituted by nodes and links between nodes (edges). Nodes represent the elements in a complex system, and the edges represent the interaction between system elements. E-commerce market is a complex network, and its complexity reflects in the following areas:

- The system has a huge number of nodes, and its network structure is complex and presents a variety of different characteristics. There are the generation and disappearance of nodes or edges. The emergence and demise of elements that have the life-cycle characteristic is very common. The relationships among elements are also dynamic changed.
- Networks are no governance structures. There is not one dominant organization to control and organize the other members in the network. Rather, the network concept is a way to visualize and understand the way firms and organizations are interconnected directly and indirectly through relationships. Networks are not under the control of individual nodes.
- Networks are formed in a self-organizing way through the actions and interactions of actions and interactions of actors involved, as they occur over time. They are continually being made and remade (or not) through ongoing structuring and restructuring processes. The multiple interactions and feedback effects continually taking place in networks lead to a complexity that makes it very difficult to control and predict for any individual actor.
- Time plays a central role in explaining and understanding exchange. Business relations develop over time and they are path-dependent. Buyers and sellers actively take into account what has happened before and they also form plans and have expectations of what is likely to happen in the future, both of which affect their decisions in the present. The state of the network subject (person or organization) changes over time.

From a mathematical point of view, network can be described by set $N=\{V, E\}$ composed with nodes set $V=\{v_1, v_2, \ldots, v_n\}$ and edges set $E=\{e_1, e_2, \ldots, e_m\}$. Network analysis framework for e-commerce market is as the following Figure 2. From the graph we can see that, there are three steps for analyzing on e-commerce market from a network point of view—definition of network, analysis of network topology and analysis of network environment. Within this analysis, a network community detection is often being performed. A community is a subgraph with many edges inside and relatively few edges outside. It clarifies relationships among vertices and the overall structure of the network. For unipartite networks composed of only one vertex type, Girvan-Newman modularity [12] is proposed for evaluating network divisions, and it is widely used for the research of community detection. Searching for the divisions with high modularity values, which is called modularity optimization, is also investigated by many researchers.

2.1 The design of our model

In this paper, we will model the e-commerce processes with the existence price comparison sites as a tripartite network. In general, tripartite networks are the networks whose vertices are composed of three disjoint sets. Nodes in this tripartite networks are basically different in each layer (see Figure 2). Here, the set X are buyers, the set Y are price comparison sites and nodes in the set Z present e-commerce sites. The reason we limit our model in this manner is to simplify the problems. In our model, we will look for communities in the different parts of the tripartite network and identify the properties of these communities.

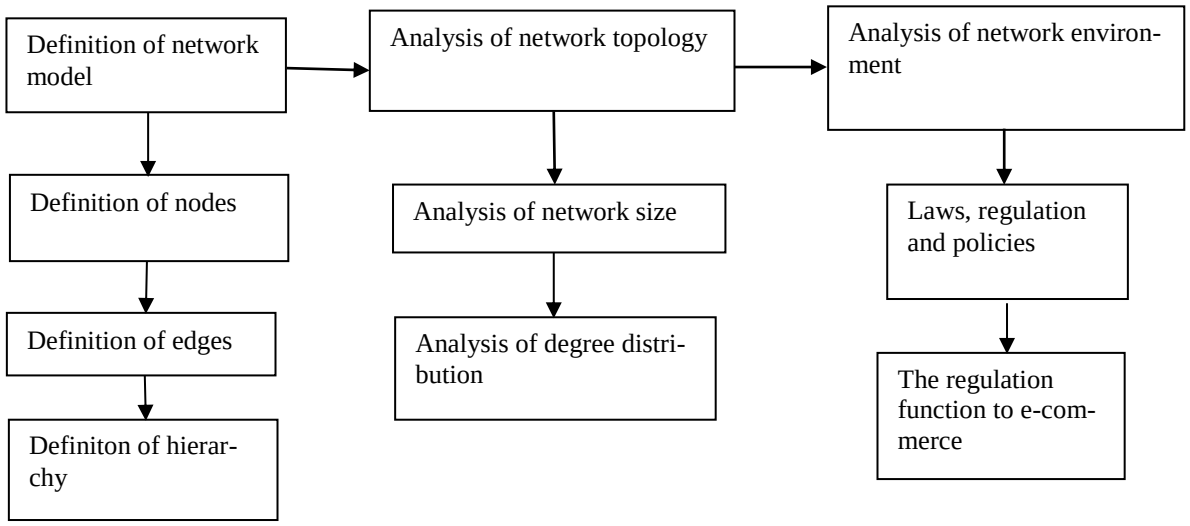


Figure 2 E-commerce network analysis framework

To detect communities from tripartite networks, they are often projected into simpler unipartite networks and then processed with conventional methods. In order to analyze multipartite networks, it is very common to study a one-mode mapping of this original network. This approach is called one-mode projection [6, 26]. The multipartite network is projected onto a one-mode network by dropping one of the two nodes sets and connecting two nodes in the one-mode network if they share a neighbor in the multipartite network. This popular approach is necessitated by the fact that many network measures cannot be directly applied to multipartite networks [16]. However, projection will lose the in-formation that original tripartite networks have [9]. As a way for solving this problem, authors [1, 19] propose tripartite modularities that evaluate the qualities of the divisions of tripartite networks. With these tripartite modularities, Murata and Ikematsu [14, 17] detect communities from tripartite networks. But these methods still have some problems such as requiring a parameter, projection and computational cost. Definition of a community in n-partite networks is controversial. Barber [1] defines it as a set of vertices of arbitrary vertex types.

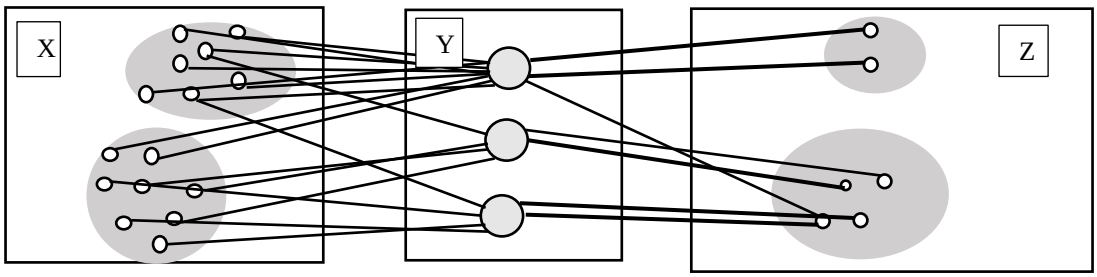


Figure 2 Examples of a tripartite network and a community structure

In our case, nodes in tripartite networks are called *X* (buyers), *Y* (price comparison sites) and *Z* (e-commerce sites), (see Figure 2). M is the number of all edges and *V* is a set of all vertices in the tripartite network. V^X is a set of vertices that belongs to vertex set X. V_l^X is a set of vertices that belongs to vertex set *X* and also to communty *l*. C^X is a set of communities in vertex set *X*. An adjacent matrix of a tripartite network is *A*, and its (*i*, *j*, *k*)-th element is represented as *A*(*i*, *j*, *k*), where $i \in V^X$, j $\in V^Y$ and k $\in V^Z$. It takes three arguments because each "hyperedge" always connects the three vertices from each of $i \in V^X$, $j \in V^Y$ and $k \in V^Z$. *A*(*i*, *j*, *k*) takes 1 when vertices *i*, *j* and k are connected with a hyperedge, and otherwise it takes 0.

In this paper, we use the approach for community detection which is based on the Girvan-Newmann algorithm [12], which is also a hierarchical method with modularity measure used to detect communities in complex systems. Modularity is one measure of the structure of networks or graphs. It was designed to measure the strength of division of a network into modules (also called groups, clusters or communities).

Girvan-Newman modularity is not appropriate for evaluating divisions of tripartite networks. Because it evaluates the density of edges within each community, but in the case of tripartite networks, there is no edge that connects the vertices of the same type. If a tripartite network is projected into unipartite or bipartite networks, Girvan-Newman modularity or other bipartite modularity can be applied [8, 19]. But the projection will lose the information that original tripartite network has [17, 18]. Murata extended Girvan-Newman modularity for tripartite networks so that no projection is required [17]. We have used the original Neubauer's form of tripartite network modularity [19] which evaluates the density of connections between communities of different vertex types, not within communities:

$$Q_N = \sum_l \sum_m \sum_n Q_{lmn} = \sum_l \sum_m \sum_n \alpha_{lmn}(e_{lmn} - a_l^X a_m^Y a_n^Z) \,. \tag{1}$$

$$\alpha_{lmn} = \frac{1}{3}\left(\frac{e_{lmn}}{a_l^X} + \frac{e_{lmn}}{a_m^Y} + \frac{e_{lmn}}{a_n^Z}\right), \tag{2}$$

where:

$$e_{lmn} = \frac{1}{M}\sum_{i \in V_l^X} \sum_{j \in V_m^Y} \sum_{k \in V_n^Z} A(i,j,k) \,, \tag{3}$$

$$a_l^X = \sum_{m \in C^Y} \sum_{n \in C^Z} e_{lmn} = \frac{1}{M}\sum_{i \in V_l^X} \sum_{j \in V^Y} \sum_{k \in V^Z} A(i,j,k) \,, \tag{4}$$

$$a_m^Y = \sum_{m \in C^X} \sum_{n \in C^Z} e_{lmn} = \frac{1}{M}\sum_{i \in V^X} \sum_{j \in V_m^Y} \sum_{k \in V^Z} A(i,j,k) \,, \tag{5}$$

$$a_n^Z = \sum_{m \in C^X} \sum_{n \in C^Y} e_{lmn} = \frac{1}{M}\sum_{i \in V^X} \sum_{j \in V^Y} \sum_{k \in V_n^Z} A(i,j,k) \,. \tag{6}$$

The used procedure has the following steps:

1. At first, all vertices are assigned randomly to different communities. Neubauer's tripartite modularity Q_N is computed.
2. For each part of the network *X*, *Y* and *Z*, repeatedly for all possible combinations of community Q_N is computed.
3. The community pair, where the increase of Q_N is the highest, is merged.
4. Step 2 and 3 are repeated. If the value of Q_N will not increase with any merging, the established community will not be changed.
5. The values of maximum Q_N during the above processes are selected as the result of community detection.

Community search in our tripartite network is easier in this case as price comparison sites (part *Y*) are considered as individual specific communities. We have performed our experiments on PC with Core i7-2600 3.4GHz CPU, 16GB RAM, and we have used Python 2.7 with the NetworkX software package.

The statistics of detected communities are shown in Table 1. The numbers of users, price comparison sites and e-shops are 355, 5 and 124, respectively. The total numbers of vertices and edges are 2458 and 854, respectively. The Neubauer's tripartite modularity value is 0.48. The statistics of detected communities are shown in Table 1. The rows represent users, tags and web pages, respectively. The columns represent the number of communities, the average size of communities, and the average numbers of correspondences per community. The values in the last column are much bigger than 1, which means that there are many-to-many correspondences among communities [18].

	Number of communities	**Average size of communities**	**Average number of correspondences per community**
User	160	2.2	5.14
E-commerce sites	40	3.1	4.85

Table 1 Statistics of detected communities (parts *X* and *Z* in Figure 2)

3 Results and discussion

We performed experiments with synthetic tripartite networks. These networks had known community structures. Synthetic networks in our experiments have scale free properties just like many real networks. The graph model

described above provides simple expressions for properties such as cluster size, but its use of random edge placement neglects some significant aspects of social networks. In particular, random edges do not correctly account for local clustering of social networks: two nodes linked to a third are significantly more likely to have a link between them than expected from random edge choices [20]. Thus, it is important to evaluate inference capabilities using more realistic networks. One approach involves graph models incorporating additional aspects of social interactions such as transitivity [22].

The random graph model allows estimating the likely extent of the largest cluster based on observed preference correlations among neighbors of a few people who express interest in a resolve. This model is less informative for distinguishing cluster size as the largest component size shows a gradual increase with preference correlation. The semantics of the links in a network affect the cluster size, indicating the importance for e-commerce applications of eliciting the type of relationship a link entails, which is not made explicit on most social network sites [13]. Thus, depending on the application, a vendor could select a type of link more or less likely to give large clusters of consumers with similar preferences for particular products or bundles. Large clusters could facilitate word-of-mouth marketing while small clusters could be useful for surveys to sample a variety of opinions less likely to influence each other. The latter is particularly relevant in using social networks since such sites often promote users learning about others' activities and preferences via their links in the network.

One may also wonder if the degree of a bottom vertex (in X part) in the tripartite graph and the classical version of the same graph are related. The degree of a vertex in the classical graph is the sum of the degrees of the top vertices to which it is connected in the tripartite graph, minus the number of vertices in common in the neighborhood of these vertices. One can easily be convinced that this overlap between neighborhoods, if any, can have a great impact on the degree distribution. To deepen this notion of overlap, one can observe the correlation between the bottom vertex degrees in both tripartite and classical version of the same graph.

4 Conclusions

We have explored a modularity appropriate for tripartite networks using edge clustering. Experimental results on synthetic networks show accuracies of the proposed method. Since the used method is not deterministic, detailed analysis of detected communities, such as comparing them or visualizing them, are left for our future work. However, we found out that further speedup is needed for much larger scale tripartite networks.

A significant challenge for inference with networks is the available information is only an approximation of people's relationships and preferences. For use of social networks in e-commerce, the network information could be incomplete and out of date, that is, noisy. Thus, in practice, evaluating the usefulness of network-based inference for e-commerce requires understanding the consequence of errors in the data. Fortunately, mechanisms relying on aggregated information from social networks are somewhat robust: performance degrades gradually rather than abruptly with noise. In such cases, estimates of consumer interests based on approximate network information is beneficial compared to not using the information at all. Evaluating the amount of noise in online networks and its effects on mechanisms relying on those networks is an important direction for future work.

A further challenge arises from the using real networks data about e-commerce. While available online networks can include thousands or millions of users, and thus give strong statistical correlations, detailed information on why users form links is usually lacking. Thus, it is difficult to distinguish links arising from prior similarity from influence of linked individuals creating similar preferences. In our future work, we want to explore further insight into the network of interest and make a more dynamic analysis of the network possible.

Acknowledgements

This study was supported by project grant No. IGS21C1, Faculty of Economics, University of South Bohemia Czech Republic.

References

[1] Barber, M.J.: Modularity and community detection in bipartite networks. *Physical Review* **E 76**(2007), 1–9.

[2] Beranek, L.: Sentiment Prediction on E-commerce Sites Based on Dempster-Shafer Theory. In: *Proceedings of the 34st International Conference on Mathematical Methods in Economics* (Kocourek, A. and Vavrousek, M., eds.). tech. Univ. Liberec, Liberec, 2016, 43-48.

[3] Beranek, L. and Nydl, V.: The Use of Belief Functions for the Detection of Internet Auction Fraud. In: *Proceedings of the 31st International Conference Mathematical Methods in Economics* 2013. Jihlava: College of Polytechnics Jihlava, 2013, 31 - 36.

[4] Beranek, L., Tlusty, P., Remes, R.: The Usage of Belief Functions for an Online Auction Reputation Model. In: *Proceedings of the 28th Conference on Mathematical Method in Economy* 2010. Economical faculty. University of South Bohemia, Ceské Budejovice, Czech Republic, September 2010, s. 49-54.
[5] Bianconi, G. and Barabasi, A. L.: Competition and Multicaling in Evolving Networks, *Europhysics Letters*, **54**(2001), 4, 436-442.
[6] Borgatti, S. P. and Halgin, D. S.: Analyzing a affiliation networks. In *The Sage Handbook of Social Network Analysis* (Carrington, P. and Scott, J., eds.). The Sage handbook of social network analysis Thousand Oaks: Sage Publications, 2011, 417 - 433.
[7] Brynjolfsson, E. and Smith, M.D.: Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* **46**(2000), 563–585.
[8] Cattuto, C., Baldassarri, A., Servedio, V.D. and Loreto, V.: Emergent community structure in social tagging systems. *Advances in Complex Systems* 11(2008), 597–608.
[9] Conaldi, G., Lomi, A., and Tonellato, M.: Dynamic models of affiliation and the network structure of problem solving in an open source software project. *Organizational Research Methods* 15(2012), 385-412.
[10] Degeratu, A.M., Rangaswamy, A., Wu, J.:. Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of Research in Marketing* **17**(2001), 1, 55–78.
[11] Easley, D. and Kleinberg, J.: *Networks, Crowds, and Markets*. Cambridge: Cambridge University Press, 2010.
[12] Girvan M. and Newman M. E. J.: Community structure in social and biological networks, *Proc. Natl. Acad. Sci. USA* **99**(2002), 7821–7826.
[13] Hogg, T., Wilkinson, D.M., Szabo, G. and Brzozowski, M.: Multiple relationship types in online communities and social networks. In *Proc. of the AAAI Symposium on Social Information Processing* (K. Lerman et al., eds.), 2008, 30 - 35.
[14] Ikematsu, K. and Murata, T.: A Fast Method for Detecting Communities from Tripartite Networks. In *Proceedings of 5th International Conference on Social Informatics* (Jatowt A., eds.), Kyoto, Japan, 2013, 192 - 205.
[15] Lamanauskas, V., Slekiene, V., Balog, A., Pribeanu, C.: Exploring the Usefulness of Social Networking Websites: a Multidimensional Model. *Studies in Informatics and Control* **22**(2013), 2, 175-184.
[16] Latapy, M., Magnien, C., and Vecchio, N. D.: Basic notions for the analysis of large two-mode networks. *Social Networks* **30**(2008), 31-48.
[17] Murata, T.: Modularity for heterogeneous networks. In: *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia* (Chignell, M. et al., eds.), ACM New York, 2010, 129–134.
[18] Murata, T.: Detecting communities from social tagging networks based on tripartite modularity. In: *Workshop on Link Analysis in Heterogeneous Information Networks*, 2011, 1–4.
[19] Neubauer, N., Obermayer, K.: Towards community detection in k-partite k-uniform hypergraphs. In: *The NIPS 2009 Workshop on Analyzing Networks and Learning with Graphs*, 2009
[20] Newman M. E. J.: The structure and function of complex networks. *SIAM Review* **45**(2003), 2, 167 - 256.
[21] Rauch, J. E.: Does Network Theory Connect to the Rest of Us? A Review of Matthew O. Jackson's Social and Economic Networks. Journal of Economic Literature, **2010**, 980-986.
[22] Garry R., Snijders, T., Wang, P., Handcock, M. and Pattison, P.: Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, **29**(2007), 192 - 215.
[23] Schweitzer, F., Fagiolo, G. and Sornette, D.: Economic Networks: The New Challenges. *Science* **325**(2009), 422.
[24] Skorpil, V., Stastny, J.: Back-Propagation and K-Means Algorithms Comparison. *Proceedings of the 8th International Conference on Signal Processing*, Guilin, China, IEEE Press, 2006, 1871-1874
[25] Stencl, M., Stastny, J.: Neural network learning algorithms comparison on numerical prediction of real data. *Proceedings of the16th International Conference on Soft Computing Mendel* 2010. Technical University of Brno, Brno, 2010, 280-285.
[26] Wasserman, S., and Faust, K.: *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press, 1994.

Fuzzy Discretization for Data Mining

Petr Berka[1]

Abstract. Data preprocessing or data preparation is the most time consuming and most laborious step in the whole data mining process. The reason for data preprocessing is twofold; it is necessary to select (or create) from available data characteristics relevant for given data mining task, and to represent these characteristics in a form suitable for selected data mining algorithm. Among the typical data operations performed in this step discretization of numeric attributes plays an important role as algorithms for creating either association or decision rules cannot handle numeric attributes directly.
Different approaches to discretization can be used. Equidistant or equifrequent discretization are typical examples of so-called "class-blind" methods since they deal only with the discretized attribute. Another group of discretization are "class-sensitive" methods; the fact that the examples (objects) belong to different classes is taken into account here. The discretization procedures typically generate sharp boundaries (thresholds) between intervals.
The paper describes a class-sensitive discretization method where the boundary between intervals is defined using fuzzy membership function. The paper shows an experimental evaluation of the proposed method on some benchmark data and also compares the proposed method with more standard (i.e. crisp) discretization. A naïve Bayesian classifier as well as a tree learning algorithm are used in the experiments.

Keywords: data mining, discretization, decision rules, fuzzy intervals.

JEL Classification: C38
AMS Classification: 62C86, 62H30

1 Introduction

Data preprocessing or data preparation is the most time consuming and most laborious step in the whole data mining process. The reason for data preprocessing is twofold; it is necessary to select (or create) from available data characteristics relevant for given data mining task, and to represent these characteristics in a form suitable for selected data mining algorithm. Among the typical data operations performed in this step discretization of numeric attributes plays an important role. The discretization process turns quantitative data into qualitative data by transforming numeric attributes into categorical ones. Discretization is necessary to perform for data mining algorithms that cannot handle numeric attributes directly (typically algorithms for creating either association or decision rules). But discretization can be also understood as a data reduction method since it maps data from a huge range of numeric values to a significantly smaller number of discrete values (intervals). From the latter point-of-view, discretization can increase the so-called class noise, i.e. situations where examples with same values of input attributes (and thus undistinguishable by any data mining algorithm) belong to different classes, but can improve the understandability of the classification models (rules, trees or Bayesian classifiers can be better interpreted by domain experts).

Different approaches to discretization can be used. The basic distinction between them is whether they are unsupervised (class blind), dealing only with the numeric attribute itself) or supervised (class sensitive) taking into account the fact that the examples (objects) belong to different classes. Further differences between the discretization algorithms are in:

- integration with machine learning algorithms (integrated or stand-alone as preprocessing tool);
- search strategy (top-down by splitting intervals or bottom-up by merging intervals);
- the impurity measure for evaluating potential intervals (entropy, information gain, $\chi 2$ test, minimum classification error);
- number of intervals (binarization or creating more intervals);
- stopping criterion (number of intervals, frequency of intervals, impurity of intervals).

A nice review of the differences between the discretization algorithms can be found in [5].

[1] University of Economics, Prague, Dept. of Information and Knowledge Engineering, W. Churchill Sq. 4, Prague 3, 130 67, berka@vse.cz.

The rest of the paper is organized as follows: section 2 reviews most typical discretization methods, section 3 presents the fuzzy discretization algorithm, section 4 shows its empirical evaluation and section 5 concludes the paper.

2 Related work

During discretization, the boundaries $LBound_i$, $UBound_i$ of intervals Int_i should be found that will cover the range of the numeric attribute. In equidistant discretization, the boundaries can be defined as

$$LBound_i = A_{min} + d * i$$
$$UBound_i = LBound_i + d$$
$$(LBound_{i+1} = UBound_i)$$

where A_{min} is the minimal value of attribute A and d is the width of the interval defined as $(A_{max} - A_{min})/k$, where k is the user-given number of intervals. In equifrequent discretization the boundaries can be defined as

$$LBound_i = (f * i)^{\text{th}}. \text{ value}$$
$$UBound_i = = (f * (i+1))^{\text{th}}. \text{ value}$$
$$(LBound_{i+1} = UBound_i)$$

where $f = (n/k)$ is the average frequency of an interval, n is the number of examples and k is the user-given number of intervals. These are the typical examples of unsupervised discretization. As an example of supervised discretization, let us consider the entropy based binarization as used in tree learning algorithm C4.5 [9]. Here, the cut-point θ being *UBound* of the first interval and *LBound* of the second interval is found in such a way, that the value

$$H(A(\theta)) = \frac{n(A(<\theta))}{n} H(A(<\theta)) + \frac{n(A(>\theta))}{n} H(A(>\theta))$$

is the smallest one for all possible settings of θ_x selected as a value $(a_{k+1} - a_k)/2$ computed for two subsequent values a_k, a_{k+1} of the attribute A. Here n is the number of examples, $n(A(<\theta))$ resp. $n(A(>\theta))$ is the number of examples having the value of attribute A smaller resp. greater than θ, and entropy for C classes, $H(A(<\theta))$ resp. $H(A(>\theta))$ is defined as $-\sum p_j \log p_j$, where $j = 1,\ldots C$ and p_j is the relative number of examples belonging to class j that have the value of attribute A smaller resp. greater than θ.

3 Fuzzy discretization

There are two good reasons to consider fuzzy discretization instead of crisp one:

1. Numeric attributes are usually continuous in their nature. This means in the context of data mining tasks like classification or prediction, that small change in the value of such attribute should not result in an abrupt change of its interpretation (classification or prediction result),
2. Because the training set is finite (even if very large), the observed list of values of a numeric attribute is finite as well thus creating "gaps" in the sequence of consecutive values. The standard way how to handle such a gap on the interval boundary is to assign the whole gap to one of the intervals (this is e.g. the case of on-line discretization in C4.5 algorithm as implemented in weka) or to place the interval boundary into the middle of the gap following the maximal margin hyperplane principle of SVM. Again, fuzzy boundaries between adjacent intervals can better handle this situation.

3.1 The algorithm

The discretization algorithm was motivated by our rule-learning algorithm KEX. KEX performs symbolic empirical multiple concept learning from examples, where the induced concept description is represented as weighted decision rules in the form

$$Ant \Rightarrow C \ (weight)$$

where *Ant* is a combination (conjunction) of attribute-value pairs,
C is a single category (class),
weight from the interval [0; 1] expresses the uncertainty of the rule.
Interested reader should refer to [1] for more details on the KEX algorithm.

We discretize each numerical attribute separately. The basic idea is to create intervals for which the aposteriori distribution of classes $P(C \mid Interval)$ significantly differs from the apriori distribution of classes $P(C)$ in the whole training data. This can be achieved by simply merging such values, for which most objects belong to the same class. Within the KEX knowledge acquisition approach, this will lead to rules of the form

$$interval \Rightarrow C,$$

but this approach can be used for other learning algorithms, too.

The discretization algorithm is shown in Fig. 1. The algorithm performs a bottom-up search by merging adjacent intervals that share similar class distribution. The process starts by considering each value a_i of the discretized attribute an initial interval. A class label is then assigned to each initial interval using the procedure ASSIGN. A special class label "UNKNOWN" is used for situations, where the distribution of examples into classes within initial interval does not significantly differ from the apriori distribution of examples into classes in the whole training set. The procedure FUZZY-INTERVAL then merges adjacent intervals that have the same class label (step 3.1). The intervals with the "UNKNOWN" label are then eventually treated as fuzzy boundaries of intervals with "regular" class label (step 3.2). A trapezoidal membership functions are used to create the fuzzy intervals.

The number of resulting intervals is controlled by specifying a threshold for minimal number of objects within one interval, and in step 3.1 by assigning the label "UNKNOWN" to less frequent intervals.

MAIN LOOP:

1. sort the values a_i of the numeric attribute in ascending order;
2. **for each** value a_i **do**
   2.1. $LBound_i := (a_{i-1}+a_i)/2$, $UBound_i := (a_i+a_{i+1})/2$
   2.2. count the frequencies of each class and store the max frequency into *maxfreq*;
   2.3. assign class label using procedure ASSIGN;
3. create intervals using procedure FUZZY-INTERVAL;

ASSIGN:

if for the given value a_i all objects belong to same class,
then assign the value to that class
else if for the given value the distribution of objects with respect to class membership significantly differs (according to χ^2 goodness-of-fit test) from frequencies of goal classes,
then assign that value to the most frequent class
else assign the value to the class "UNKNOWN";

FUZZY-INTERVAL:

3.1. **if** for sequence of values the majority class is the same,
then create the interval $INT_i = [LBound_i, UBound_i]$ from these values (with characteristic function set to 1 in the entire range $[LBound_i, UBound_i]$;

3.2. **if** the interval INT_i belongs to the class "UNKNOWN"
then
if its neighbouring intervals INT_{i-1} , INT_{i+1} belong to the same class
then create the interval by joining $INT_{i-1} \cup INT_i \cup INT_{i+1}$ with characteristic function set to 1 in the whole range $[LBound_{i-1}, UBound_{i+1}]$;
else create one interval by joining $INT_{i-1} \cup INT_i$ with characteristic function set to 1 in the range $[LBound_{i-1}, UBound_{i-1}]$ and set to $\frac{x - LBound_i}{UBound_i - LBound_i}$ for x in the range $[LBound_i, UBound_i]$, and second interval by joining $INT_i \cup INT_{i+1}$ with characteristic function set to 1 in the range $[LBound_{i+1}, UBound_{i+1}]$ and set to $\frac{LBound_i - x}{UBound_i - LBound_i}$ for x in the range $[LBound_i, UBound_i]$;

3.3. create continuous coverage of the attribute by treating „gaps" between intervals as intervals of class "UNKNOWN" (in the same way as in the step 3.2)

Figure 1 Fuzzy discretization algorithm

When discretizing values of one numerical attribute of an object using our fuzzy discretization procedure, we can obtain up to two categories (discretized values) with the sum of membership functions equal to 1. Thus, the original example with one numerical attribute can be split (according to the value of this attribute) into two examples. Let us call such examples partial. This splitting of objects can be repeated for every numerical attribute. The weight $w(\mathbf{x}_i)$ of a partial example $\mathbf{x}_i$ is

$$w(\mathbf{x}_i) = \prod_j \mu_{A_j}(x_{ij}),$$

where $\mu_{A_j}(x_{ij})$ is the degree of membership of value x_{ij} to the interval A_j.

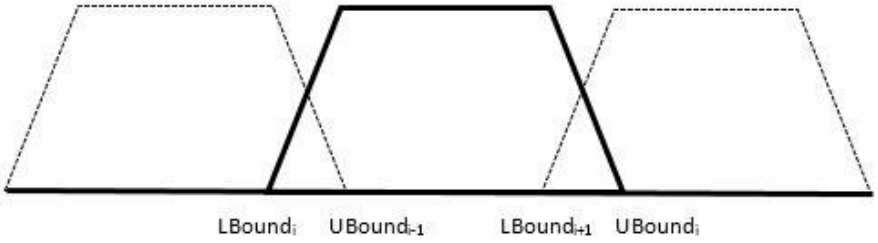


Figure 2 Fuzzy intervals

Let us assume, that the initial data contain an example

[... hair(black), temp(36.9), BMI(25.2), smoker(yes)...].

If during the value temp(36.9), will be assigned to the interval temp(low) with $\mu(x) = 0.8$ and to the interval temp(high) with $\mu(x) = 0.2$ and if the value BMI(25.4) will be assigned to the interval BMI(normal) with $\mu(x) = 0.3$ and to the interval BMI(overweight) with $\mu(x) = 0.7$, then the original example will be split into following four examples:

[... hair(black), temp(low), BMI(normal), smoker(yes)...]: w(x) = 0.8 x 0.3 = 0.24
[….hair(black), temp(high), BMI(normal), smoker(yes)...]: w(x) = 0.2 x 0.3 = 0.06
[... hair(black), temp(low), BMI(overweight), smoker(yes)...]: w(x) = 0.8 x 0.7 = 0.56
[... hair(black), temp(high), BMI(overweight), smoker(yes)...]: w(x) = 0.2 x 0.7 = 0.14

We can then use all such examples or only the example with the highest weight $w(\mathbf{x})$ as the result of the discretization; this gives us two possible modes of discretization.

4 Empirical evaluation

We will demonstrate the functionality of the proposed fuzzy discretization on some benchmark data taken from the UCI Machine Learning Repository [9]. The Australian Credit data set concerns credit card applications. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. The Pima Indians Diabetes data concerns a medical screening of the Indians population living near Phoenix, Arizona, USA. The diagnostic, binary-valued variable investigated is whether the patient shows signs of diabetes according to World Health Organization. The iris data set is a famous data set prepared by R. A. Fisher in 1936. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. The Japanese Credit Screening database contains examples representing positive and negative instances of people who were and were not granted credit, respectively. Basic characteristics of the used data are shown in Tab. 1.

We run the fuzzy discretization algorithm in both modes. All partial examples created by discretizing an original example are used in the FULL mode, only the partial example with the highest weight w is used in the MAX mode. To evaluate the discretization algorithm more thoroughly, we also use a crisp discretization and the original data in the experiments. The crisp discretization differs from the fuzzy one in the function INTERVAL. The CRISP-INTERVAL function shown in Fig. 3 is used instead of the FUZZY-INTERVAL function of the discretization algorithm shown in Fig. 1.

Dataset	no. objects	no. input attributes (numeric/categorical)	no. classes	freq. majority class	default accuracy
Australian credit	690	6/8	2	56%	100%
Diabetes	768	8/0	2	65%	100%
Iris	150	4/0	3	33%	100%
Japan credit	125	5/5	2	68%	100%

Table 1 Characteristics of the used data

The effect of discretization was evaluated indirectly, by running a particular machine learning algorithms. As the experiments were conducted on both numeric and categorical data, we need to use an algorithm that can handle both types of data directly. We thus choose a decision tree learning algorithm. We run the experiments using Rapid Miner, one of the most popular freely available data mining system. Rapid Miner can directly work with partial examples, so no modifications to the tree learning algorithm were necessary and the experiments were very straightforward. Tables 2 and 3 show the results of the experiments in the terms of overall classification accuracy (the number of correctly classified examples divided by the number of all examples) obtained from 10-fold cross-validation for the decision tree algorithm and naïve Bayes classifier respectively.

The experiments on the benchmark data do not show a clear picture but they seem to indicate that the discretization has a positive effect, i.e. it can improve the classification accuracy of both decision tree and naïve Bayesian classifier. When looking closely at the numbers in Tables 2 and 3, the max fuzzy strategy was the "winner" in most of the experiments.

CRISP-INTERVAL:

3.1. **if** a sequence of values belongs to the same class,
then create the interval $INT_i = [LBound_i, UBound_i]$ from these values (with characteristic function set to 1 in the entire range $[LBound_i, UBound_i]$;

3.2. **if** the interval INT_i belongs to the class "UNKNOWN"
then
if its neighbouring intervals INT_{i-1} , INT_{i+1} belong to the same class
then create the interval by joining $INT_{i-1} \cup INT_i \cup INT_{i+1} = [LBound_{i-1}, UBound_{i+1}]$;
else create the interval either by joining $INT_{i-1} \cup INT_i = [LBound_{i-1}, UBound_i]$ or by joining $INT_i \cup INT_{i+1} = [LBound_i, UBound_{i+1}]$ according to higher value of χ^2;

3.3. create continuous coverage of the attribute by by assigning $LBound_i := (LBound_i + UBound_{i-1})/2$ and $UBound_{i-1} := LBound_i$

Figure 3 Creating crisp intervals

	Decision Tree Classification accuracy			
Data set	**original**	**crisp**	**full fuzzy**	**max fuzzy**
Australian Credit	84.8%	84.2%	87.3%	80.9%
Diabetes	70.6%	75.0%	72.5%	77.0%
Iris	65.3%	96.0%	94.9%	98.3%
Japan credit	72.8%	70.2%	84.5%	72.7%

Table 2 Classification accuracy for decision tree algorithm

	Naïve Bayes Classification accuracy			
Data set	**original**	**crisp**	**full fuzzy**	**max fuzzy**
Australian Credit	64.1%	85.2%	86.6%	53.5%
Diabetes	75.3%	78.3%	78.4%	79.5%
Iris	95.3%	93.3%	94.7%	98.1%
Japan credit	75.2%	80.0%	80.2%	81.4%

Table 3 Classification accuracy for naïve Bayes classifier

5 Conclusions

The paper describes a class-sensitive discretization method where the boundary between intervals is defined using fuzzy membership function. The discretized values thus form a so-called fuzzy partition of the universal set of the attribute.

We propose two strategies of fuzzy discretization of numeric attributes and compare them with a standard discretization as well as with the original data. The comparison is based on running classifier algorithms (naïve Bayesian classifier and decision tree) and comparing their classification accuracies. As we run both classifiers using the same (default) settings in all the experiments, the differences in classification accuracies are caused only by different forms of discretization.

Acknowledgements

The paper was processed with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Berka, P.: Learning compositional decision rules using the KEX algorithm. *Intelligent Data Analysis*, Vol. 16, No. 4 (2012), 665-681.
[2] Berka, P. and Bruha, I.: Empirical Comparison of Various Discretization Procedures. *Int. J. of Pattern Recognition and Artificial Intelligence*, Vol. 12, No. 7, (1998), 1017-1032.
[3] Elomaa, T. and Rousu, J.: Fast Minimum Error Discretization. In: *Proceedings. of 19th Int. Conf. On Machine Learning (ICML2002)*, Morgan Kaufman, 2002, 131-138.
[4] Fayyad, U.M. and Irani, K.B.: Multi-interval Discretization of Continuous-Valued Attributes for Classification Learning. In: *Proceedings of 13th Int. Joint Conf. of Artificial Intelligence (IJCAI 1993)*, 1993, 1022-1029.
[5] Garcia, S., Luengo, J., Saez, J.A., Lopez, V. and Herrera, F.: A Survey of Discretization Techniques: Taxonomy and Empirical Analysis in Supervised Learning. *IEEE Trans. On Knowledge and Data Engineering*, Vol. 25, No. 4 (2013), 734-750.
[6] Choi, Y-S. and Moon, B.R.: Feature Selection in Genetic Fuzzy Discretization for the Pattern Classification Problems. *IEICE Trans. Inf. & Syst*., Vol. E90-D, No. 7 (2007), 1047-1054.
[7] Kerber, R.: ChiMerge: Discretization of Numeric Attributes. In: *Proceedings National Conf. on Artificial Intelligence (AAAI 1992)*, 1992, 123-128.
[8] Lee, C and Shin, D.: A context sensitive discretization of numeric attributes for classification learning. In: *Proceedings of European Conf. on Artificial Intelligence (ECAI-94)*, John Wiley, 1994, 428 432.
[9] Lichman, M.: UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. 2013.
[10] Pyle, D.: *Data Preparation for Data Mining*. Morgan Kaufmann Publ., 1999.
[11] Quinlan, J.R.: C4.5: *Programs for machine learning*. Morgan Kaufmann Publ., 1994.

Wages in the Czech Regions: Comparison and Wage Distribution Models

Diana Bílková[1]

Abstract. This paper deals with the situation regarding wages in the individual regions of the Czech Republic. Attention is paid to the context of employees' average wage with other economic indicators, such as GDP per capita, unemployment rate, average old-age pension, economic activity rate, average residential price and average rent for apartment 3 + kitchen, all by individual regions. Cluster analysis is used to obtain the regions similar to one another in terms of these variables. Methods of the furthest neighbor, Euclidean distance metric and classification of the regions of the Czech Republic into three, five or seven clusters are used in this cluster analysis. Wage distribution models are constructed and mutually compared. Three-parametric lognormal curves represent the basis of the theoretical distribution and parameters of the distribution are estimated using maximum likelihood method. The accuracy of the obtained models has been evaluated using Akaike and Bayesian information criterions. The prediction of wage level for individual regions for next two years is constructed using the time series analysis and exponential smoothing. Data concerning the above indicators come from the official website of the Czech Statistical Office and they cover the period since 2009.

Keywords: Cluster analysis, method of the furthest neighbor, Euclidean distance metric, three-parametric lognormal curves, maximum likelihood method, modeling of wage distribution by regions, prediction of wage level by regions, time series analysis, exponential smoothing.

JEL Classification: C02, C10, C55, C38, C46, C13, C22

AMS Classification: 62H30, 60E05, 62H12, 62M10

1 Introduction

The main aim of this paper is to create the regions similar to one another in terms of employees' average wage and other economic indicators, such as GDP per capita, general unemployment rate, average old-age pension, economic activity rate, average residential price and average rent for apartment 3 + kitchen in 2015. The main variable is gross monthly wage. There are annual data related to gross monthly wages in the year. For example, average wage then represents average gross monthly wage during the year. Data for this research come from the official website of the Czech Statistical Office (CSO).

The importance of clustering of the Czech regions according to the level of wages in the Czech regions lies in providing information for the Government of the Czech Republic on the labor market situation in individual regions of the Czech Republic. One of the causes of low wage level in the region is often high unemployment rate in this region. This analysis may be an incentive for the Government of the Czech Republic about various socio-economic measures and measures to promote employment growth and growth of wage level in the regions representing the clusters with the lowest wage level in order to narrow the gap between regions.

Cluster analysis was used to allocate the Czech regions into relatively homogenous groups according to the mentioned economic indicators in these regions. Methods of the furthest neighbor, Euclidean distance metric and classification of the regions of the Czech Republic into three, five or seven clusters are used in this cluster analysis. Multivariate statistical data analysis, which are often used for processing economic data (see for example [8]), may also include other multivariate methods of statistical data analysis, namely canonical correlation analysis. For example, [7] or [9] deal with the special aspects of cluster analysis, [11] uses cluster analysis for clusters of EU's regions according to demographic criteria. Cluster analysis is described in detail in [2].

Construction of wage distribution models for individual regions is no less important aim of this paper. The development of wage distribution during 2009–2015 is watched. The issue of various approaches for the acquisition of robust parameter estimations of continuous probability distribution is quite widely solved in the statistical literature now; see for example [10]. Three-parametric lognormal curves (see for example in [4] or [6]) represent

[1] University of Economics, Prague; Faculty of Informatics and Statistics; Department of Statistics and Probability; Sq. W. Churchill 1938/4; 130 67 Prague 3; Czech Republic; e-mail: bilkova@vse.cz.

the basic theoretical probability distribution and parameters of the distribution are estimated using maximum likelihood method. The substance of this method is explained in detail in [3] or [5]. The next aim of this paper represents the prediction of wage level for individual regions for next two years and it is constructed using the time series analysis, see [1].

2 Results

Table 1 presents an official designation of the individual regions in the Czech Republic, which is derived from the name of the appropriate county town.

Bohemia				Moravia	
Region	**Indication**	**Region**	**Indication**	**Region**	**Indication**
Capital Prague Region	A	Usti Region	U	South Moravian Region	B
Central Bohemian Region	S	Liberec Region	L	Olomouc Region	O
South Bohemian Region	C	Hradec Kralove Region	H	Zlin Region	Z
Pilsen Region	P	Pardubice Region	E	Moravian Silesian Region	T
Karlovy Vary Region	K	Vysocina Region	J		

Table 1 Official designation of the individual regions in the Czech Republic

Number of classes					
Three clusters		**Five clusters**		**Seven Clusters**	
1st cluster	A	1st cluster	A	1st cluster	A
2nd cluster	S	2nd cluster	S	2nd cluster	S
	P		P		P
	B		B	3rd cluster	C
	E		E		L
	Z		Z		H
3rd cluster	C	3rd cluster	C		J
	K		L		O
	U		H	4st cluster	K
	L		J	5st cluster	U
	H		O		T
	J	4st cluster	K	6th cluster	E
	O	5th cluster	U		Z
	T		T	7th cluster	B

Table 2 Clusters of Czech regions – three, five and seven clusters

Table 2 provides an overview of the results of cluster analysis of the Czech regions (into three, five or seven clusters) according to the wage level, GDP per capita, general unemployment rate, average old-age pension, economic activity rate, average residential price and average rent for apartment 3 + kitchen. Figure 1 shows average gross monthly wage in individual Czech regions in 2015. It is clear from Table 2 that the first cluster has only one element (Capital Prague Region) in all cases. It follows from Figure 1 that it is region with the highest wage level in the Czech Republic.

In the case of classification of the Czech regions into three clusters, five regions represent the second cluster. There are the regions with higher wage level, which are Central Bohemian Region, Pilsen Region and South Moravian Region, and regions with lower wage level, which are Pardubice Region and Zlin Region. Remaining eight regions fall into the third cluster. In the case of classification of the Czech regions into five clusters, the same five regions as in the case of classification of the Czech regions into three clusters constitute the second cluster. The third cluster has five elements, too. There are South Bohemian Region, Liberec Region, Hradec Kralove Region, Vysocina Region and Olomouc Region, so regions with middle or low wage level. The fourth cluster has only one region, which is simultaneously the region with the lowest wage level in the Czech Republic. This is Karlovy Vary Region. The fifth cluster is presented of Usti and Moravian Silesian Regions. In the case of classification of the Czech regions into seven clusters, the second cluster is presented of two regions, with the highest wage level in the Czech Republic with the exception of Capital Prague Region. There are Central Bohemian Region and Pilsen Region. The third cluster has five elements (South Bohemian Region, Liberec Region, Hradec Kralove Region, Vysocina Region and Olomouc Region), the fourth cluster is presented of only one region with the lowest wage

level in the Czech Republic (Karlovy Vary Region). The fifth cluster has two members, i.e. Usti and Moravian Silesian Regions as in the case of classification only into five clusters. Pardubice and Zlin Regions represent the sixth cluster and the seventh cluster has again one member (South Moravian Region).

This is evident from the results that people living in Prague have undoubtedly the highest standard of living as far as the material aspect is concerned. The representation of individual clusters also depends greatly on the number of clusters selected. For this reason, several cases of division of Czech regions into different number of clusters were considered. However, the results obtained can be considered to be reasonably robust in terms of the mutual similarity of the objects (regions) within each cluster. Too detailed grading reduces the clarity of the results obtained and it is therefore undesirable. There are various methods for determining the optimal number of clusters in cluster analysis. However, there is no definitive answer to the question of determining the optimal number of clusters. The problem lies in that cluster analysis is basically an exploratory approach. Interpretation of the resulting hierarchical structure depends on the context, and there are often several solutions from the theoretical point of view.

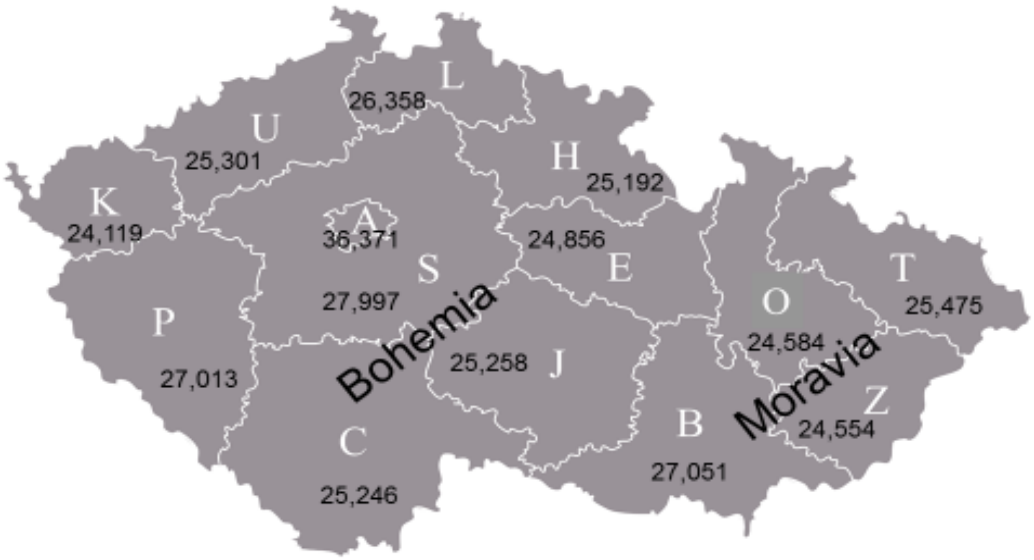


Figure 1 Placement of individual regions and average gross monthly wage (in CZK) in the Czech Republic in 2015 by regions

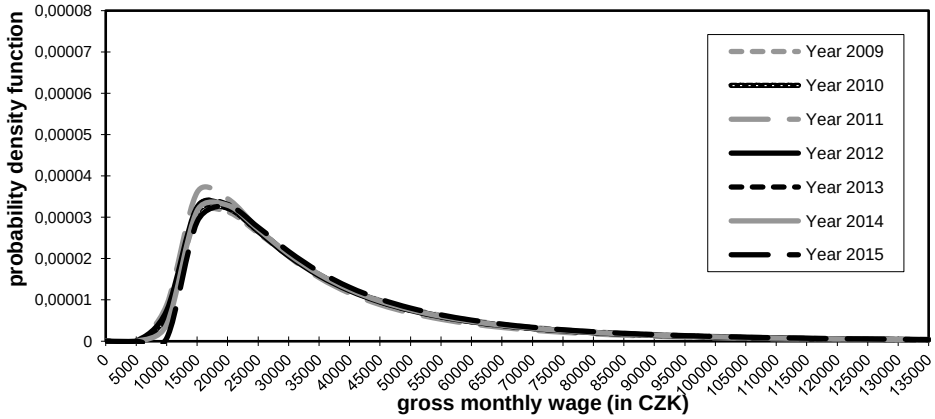


Figure 2 Development of model wage distributions − Capital Prague Region

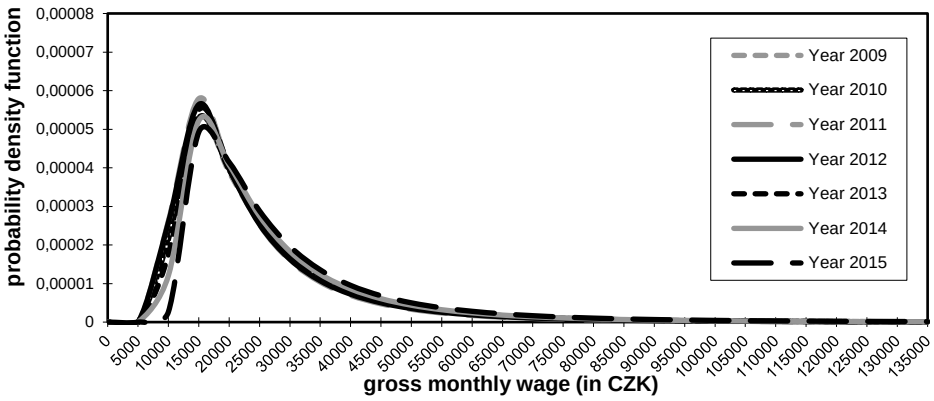


Figure 3 Development of model wage distributions − Central Bohemian Region

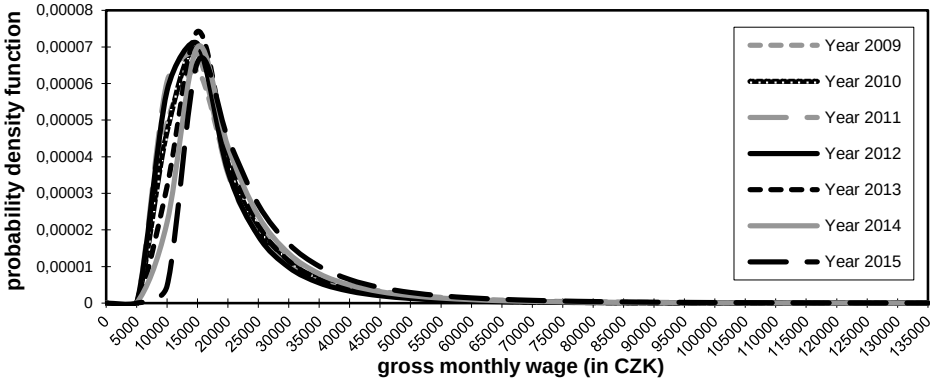


Figure 4 Development of model wage distributions − Karlovy Vary Region

		Year						
Region	**Est.**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Capital Prague Region	μ	9,937	9,911	9,803	9,885	9,875	9,885	9,907
	σ^2	0,765	0,746	0,740	0,761	0,752	0,765	0,777
Central Bohemian Region	μ	9,426	9,443	9,342	9,366	9,412	9,467	9,512
	σ^2	0,732	0,707	0,683	0,702	0,699	0,711	0,742
South Bohemian Region	μ	9,112	9,156	9,124	9,092	9,176	9,250	9,312
	σ^2	0,701	0,688	0,633	0,633	0,627	0,632	0,668
Karlovy Vary Region	μ	9,113	9,054	8,978	8,963	9,060	9,149	9,239
	σ^2	0,737	0,627	0,686	0,635	0,577	0,616	0,670
Usti Region	μ	9,264	9,280	9,130	9,169	9,216	9,266	9,337
	σ^2	0,711	0,674	0,671	0,664	0,637	0,654	0,707
Pardubice Region	μ	9,225	9,168	9,105	9,118	9,160	9,231	9,306
	σ^2	0,920	0,697	0,642	0,660	0,642	0,650	0,704
South Moravian Region	μ	9,358	9,394	9,268	9,311	9,361	9,404	9,460
	σ^2	1,028	1,025	1,028	1,027	1,018	1,014	1,009

Table 3 Estimations of parameters of three-parametric lognormal distribution using the maximum likelihood method

Theoretical wage models using three-parametric lognormal curves and maximum likelihood method of parametric estimation were constructed for selected regions. The beginning of these curves is the minimum wage in the year. Based on the results of cluster analysis with three clusters, one region from each cluster was chosen and the development of probability density functions since 2009 was captured, see figures 2–4. These figures enable some comparison of the development of wage distribution of the regions with the highest wages on the one hand and with the lowest wages on other hand during the last seven years. We can see from these figures that wage distributions with higher wage level are also distinguished by higher variability than those with lower wage level.

In addition, wage distributions with lower level of wage are more positively skewed and they have also higher kurtosis than those with higher wage level, see Figures 2–4. Based on the results of cluster analysis with seven clusters, one region from each cluster was chosen again. Table 3 shows the remaining two parameters estimated for seven selected regions. The accuracy of these models was evaluated using Akaike and Bayesian information criterions.

Region	**Exponential Smoothing**
Capital Prague Region	Holt's Linear Exponential Smoothing with $\alpha = 0.4078$ and $\beta = 0.3739$
Central Bohemian Region	Brown's Quadratic Exponential Smoothing with $\alpha = 0.9999$
South Bohemian Region	Brown's Quadratic Exponential Smoothing with $\alpha = 0.9999$
Karlovy Vary Region	Brown's Quadratic Exponential Smoothing with $\alpha = 0.9982$
Usti Region	Holt's Linear Exponential Smoothing with $\alpha = 0.8761$ and $\beta = 0.2262$
Pardubice Region	Holt's Linear Exponential Smoothing with $\alpha = 0.9999$ and $\beta = 0.1601$
South Moravian Region	Holt's Linear Exponential Smoothing with $\alpha = 0.8010$ and $\beta = 0.2262$

Table 4 Exponential Smoothing by regions

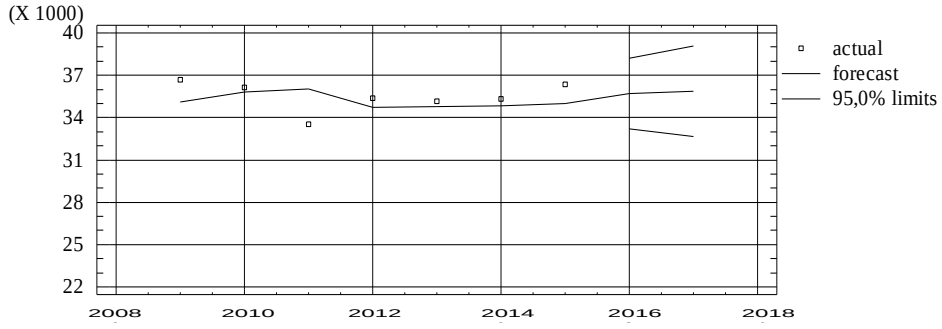


Figure 5 Holt's Linear Exponential Smoothing with $\alpha = 0.4078$ and $\beta = 0.3739$ – Capital Prague Region

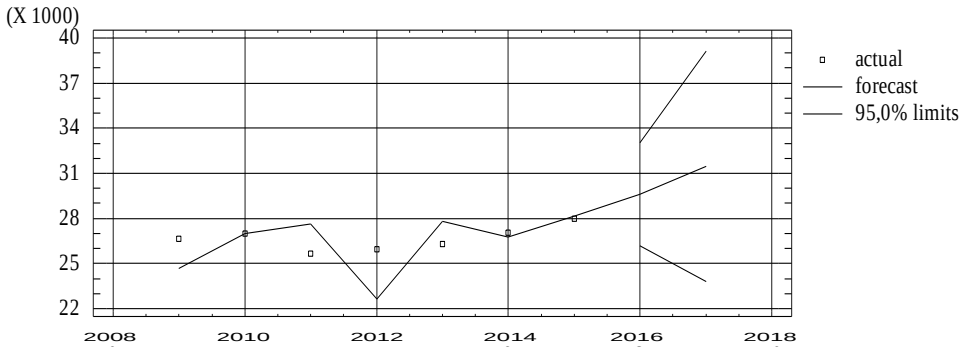


Figure 6 Brown's Quadratic Exponential Smoothing with $\alpha = 0.9999$ – Central Bohemian Region

Region	**2016**	**2017**	**Region**	**2016**	**2017**
Capital Prague Region	35,708	35,896	Usti Region	25,450	25,717
Central Bohemian Region	29,629	31,468	Pardubice Region	25,155	25,454
South Bohemian Region	26,616	28,182	South Moravian Region	27,213	27,552
Karlovy Vary Region	24,385	25,087			

Table 5 Forecasts of average gross monthly wage (in CZK) for 2016 and 2017 by regions

Table 4 and Figures 5–6 represent the methods of exponential smoothing used for predictions of average gross monthly wage for 2016 and 2017 by regions in Table 5 for seven selected regions. The suitable exponential smoothing was chosen using interpolation criteria. We can see from Table 4 that only Holt's Linear Exponential Smoothing and Brown's Quadratic Exponential Smoothing were used as the most suitable for construction of the predictions. We can expect the growth of wage level in all Czech regions for the future period, which is in agreement of awaiting of the economists.

3 Conclusion

It follows from the obtained results that Capital Prague Region has a specific position in view of other Czech Regions. This region is clearly best in terms of monitored economic indicators. People in Central Prague Region have the highest wages and people in Karlovy Vary Region have, on the contrary, the lowest wages. The average gross monthly wage of the employees in Central Prague Region was 36,371 CZK in 2015, and in the Karlovy Vary Region it was only 24,119 CZK in the same year. Dwellers of Central Bohemian region have relatively high wages, which average was 27,997 CZK in 2015. Pilsen and South Moravian Regions are characterized by relatively high wage levels, too. There are on average 27,013 CZK in Pilsen Region and 27,051 in South Moravian Region in 2015. However, counties with high wages are also at the same time characterized by relatively high gender wage gaps. It follows from the obtained time series of average gross monthly wage that we can expect the growth of wage level in all Czech regions. We can await rather quick wage growth in Central Bohemian Region, South Bohemian Region and Pilsen Region in 2016 and 2017. In other regions, including Capital Prague Region, the wage growth will be rather slow during the following two years.

The average wage currently grows in the Czech Republic. The shortage of workers is above all the wage growth. Firms are so forced to lure the employees for rising wages. The moving is one way to get a higher gross wage for the same job. Wage level very differs in the Czech Republic. The lowest wages are reported by people in the Karlovy Vary Region. On the contrary, they are the best in the Capital Prague Region, where the average wage is almost of ten thousand higher than the national average. Prague wages are seemed to belong to another, richer state. The reason for this fact can consist in the higher productivity of the local population, focusing on the progressive fields of finance and informatics. A policy of the firms may play some role, which often produce in out of Prague regions, but they add up the profits in the capital, where they have their headquarters.

When moving, it is possible to have a higher gross monthly wage for the same or similar work. At the same time, however, it is necessary to take into account the different living costs in individual regions and cities. Monthly housing expenses are particularly different. The costs of purchasing one's own flat are considerably higher in the Capital Prague Region than in other regions. Rental costs are also higher in the Capital Prague Region. When looking for a new job in another region, it is necessary to anticipate this situation in advance.

Acknowledgements

This paper was subsidized by the funds of institutional support of a long-term conceptual advancement of science and research number IP400040 at the Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic.

References

[1] Brockwell, P., and Davis, R. A.: *Introduction to Time Series and Forecasting*. Springer, New York, 2002.
[2] Everitt, B. S., Landau, S., Leese, M., and Stahl, D.: *Cluster analysis*. John Wiley & Sons, New York, 2011.
[3] Gould, W., Pitblado, J., and Poi, B.: *Maximum Likelihood Estimation with Stata*. Stata Press, Lakeway Drive, 2010.
[4] Johnson, N. L., Kotz, S., and Balakrishnan, N.: *Continuous Univariate Distributions 1*. John Wiley & Sons, New York, 1994.
[5] Johnson, N. L., Kotz, S., and Balakrishnan, N.: *Continuous Univariate Distributions 2*. John Wiley & Sons, New York, 1995.
[6] Kleiber, C., and Kotz, S.: *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons, New York, 2003.
[7] Makhalova, E., and Pecáková, I.: The Fuzzy Clustering Problems and Possible Solutions. In: *The 9th International Days of Statistics and Economics (MSED 2015)* [online]. Melandrium, Prague, 2015, 1052–1061.
[8] Malec, L.: Some Remarks on the Functional Relation Between Canonical Correlation Analysis and Partial Least Squares. *Journal of Statistical Computation and Simulation* **86** (2016), 2379–2391.
[9] Řezanková, H., and Löster, T.: Shluková analýza domácností charakterizovaných kategoriálními ukazateli (Cluster Analysis of Households Characterized by Categorical Indicators). *E+M Ekonomie a management* **16** (2013), 139–147.
[10] Sládek, V. Comparison of Robust Estimates. In: *The 16th Conference on Applied Mathematics (APLIMAT 2017)* [flash disk]. Vydavateľstvo Spektrum STU, Bratislava, 2017, 1427–1438.
[11] Šimpach, O., and Pechrová, M.: Cluster analysis of EU's regions according to demographic criteria. . In: *The 10th International Days of Statistics and Economics (MSED 2016)* [online]. Melandrium, Prague, 2016, 1806–1815.

Study Results at Faculty of Management in Jindřichův Hradec

Vladislav Bína[1], Jiří Přibil[2]

Abstract. A contemporary demographical shift in the cohort of applicants for university studies and Bachelor Degree students shows a significant change of behavior and demands a change of strategy in university marketing. In recent years we have observed in most Czech universities that applicants are required to pass lower thresholds than in the past in order to be admitted, or rather frequently even start their studies without a requirement to pass any entrance exams. Moreover, university students more frequently fail to finish their studies, or extend their study period, which is particularly noticeable in the blended learning form.
Since the numbers of students represent a basis of an important part of university funding, this paper aims to analyze the most important factors influencing study results of university students. The study data is mostly categorical; thus we search for important factors affecting the results of study and analyze dependency structure of students' achievements during their studies using the apparatus of compositional models.
The compositional models rank among the means of probabilistic modeling and present an alternative capable of representing and modeling dependency structures without the necessity to employ graphical apparatus of directed acyclic graphs (unlike the Bayesian networks confusingly hinting at non-existing causal relations).

Keywords: Faculty of Management, study results, blended learning, probabilistic dependency structure, compositional model.

JEL classification: C18,C65,I21
AMS classification: 68T30,62H20,62P25

1 A Piece of Motivation

The presented paper aims to contribute to the analysis of data concerning Bachelor Degree students studying (since 2011) in a rebuilt modularized study program. Since the issue of study problems, factors affecting probability of successful graduation and results in particular subjects are complex, we significantly simplify the analysis and focus only on determinants of study results in particular compulsory subjects. In order to simplify our considerations, we also exclude language subjects and sports, which do not appear to be problematic and are successfully finished by almost every active student.

Study results in the particular subjects are represented by categorical grades on a standard grading scale used by the University of Economics, Prague. The student receives a grade of '1' for outstanding performance with 90–100 points, a grade of '2' for a very good result of 75–89 points, a grade of '3' for good results between 61–74, a grade of '4+' for insufficient results with possibility of repetition between 50 and 59 points, and '4' for students failing with less than 50 points. In the case of state exams, students cannot obtain a '4+' grade (but in the case of failure they can repeat twice). Prerequisite courses are finished with 'Z' in the case of successful completion and by 'NZ' in the case of failure. Special category '–' is reserved for students who did not show at exams and 'O' for excused unsuccessful ending.

As we mentioned above, the studied topic is rather complex and very often concerns unobserved (or even unmeasurable) factors. Therefore, we study a simplified variant of the problem and, thanks to the categorical nature of the study data, it is natural to employ means of description used for multidimensional problems of uncertain character. These are usually handled using multivariate probabilistic tools.

The data file concerns 1,827 Bachelor Degree students and we study 24 important factors and their dependence structure. It is obvious that it is impossible to sample a multivariate distribution with 24 categorical variables. In the case of dichotomic variables it would mean to estimate more than 16 million table cells; our situation is even worse. Therefore, it is important to construct a multidimensional model which can be composed from marginals of lower dimensions. This can be done using a compositional model, which is a probabilistic model capable of

[1]University of Economics, Prague, Faculty of Management, Jarošovská 1117/II, 37701 Jindřichův Hradec, Czech Republic, bina@fm.vse.cz
[2]University of Economics, Prague, Faculty of Management, Jarošovská 1117/II, 37701 Jindřichův Hradec, Czech Republic, pribil@fm.vse.cz

representing and modeling dependency structures without the necessity to employ graphical apparatus of directed acyclic graphs (unlike the Bayesian networks confusingly hinting at non-existing causal relations). For basics of the theory of compositional models see, e.g., Jiroušek [7] and Bína and Jiroušek [5]. For the construction of compositional models we use an adaptation of statistical structure learning principles based on the likelihood ratio test of independence (see Agresti [1]) and its important property of decomposability (in context of compositional models see, e.g., Bína [3] or [4]).

2 Notions and Methodology

As sketched above, the structure learning can be based on the so-called neighborhood structures of decomposable compositional models (see [4]) and on the decomposability of a likelihood criterion usable for local computation in the process of (sub)optimal search in the space of decomposable compositional models (see [3]). Let us summarize the most important properties essential for the formulation of the structure learning algorithm.

For $K \subset N$, symbol $\kappa(x_K)$ denotes a $|K|$-dimensional distribution of variables from the system $X_K = \{X_i\}_{i\in K}$, which is defined on all subsets of a Cartesian product $\mathbf{X}_K = \times_{i\in K}\mathbf{X}_i$. In order to keep the notation simple, symbol $\kappa(x_K)$ is also used to denote a value of probability distribution κ at the point x_K. For $L \subset K$ the symbol $\kappa(x_L)$ denotes the corresponding marginal distribution.

2.1 Composition

Two multidimensional distributions can be composed in the following manner (embedding a relation of conditional independence between groups of variables).

Definition 1 (Operator of composition)**.** For two distributions $\kappa \in \Pi^{(K)}$ and $\lambda \in \Pi^{(L)}$ such that $\kappa(x_{K\cap L}) \ll \lambda(x_{K\cap L})$, their *composition* is defined by the formula

$$\kappa(x_K) \triangleright \lambda(x_L) = \kappa(x_K)\lambda(x_{L\setminus K} \,|\, x_{K\cap L}).$$

Here the symbol $\ll$ stands for the relation of *dominance* (also referred to as *absolute continuity*).

The operator of composition can be iterated and the result of the repeated application to the sequence of low-dimensional distributions is (if defined) a multidimensional distribution. The resulting multidimensional distribution $\hat{\kappa}$ is so-called *compositional model*

$$\hat{\kappa} = (\ldots((\kappa_1 \triangleright \kappa_2) \triangleright \kappa_3) \triangleright \ldots) \triangleright \kappa_n.$$

Because of its properties, this model can be written as a plain sequence of low-dimensional distributions (a *generating sequence*), where only sets of variable indices are noted $\hat{\kappa} = (K_1 \bullet K_2 \bullet \ldots \bullet K_k)_\kappa$ for the sake of simplicity.

2.2 Neighborhood Structure

The class of decomposable models corresponds to an analogous subclass of undirected graphs and, in the case of compositional models, we can also introduce its definition using the running intersection property (RIP); see, e.g., Koller and Friedman [8].

Within the class of decomposable compositional models, we can introduce a neighborhood structure given by the following pair of assertions (Theorems 1 and 2).

Theorem 1. *For any decomposable model $\hat{\kappa} = (K_1 \bullet K_2 \bullet \ldots \bullet K_k)_\kappa$ (with the exception of the independent model as a product of one-dimensional marginals) there exists a decomposable model $\hat{\kappa}' = (K'_1 \bullet K'_2 \bullet \ldots \bullet K'_{k'})_\kappa$ where one additional conditional independence relation is introduced between a pair of variables which appear (as a whole) in only one set of indices K_i ($i \in \{1, \ldots, k\}$).*

Theorem 2. *For any decomposable model $\hat{\kappa} = (K_1 \bullet K_2 \bullet \ldots \bullet K_k)_\kappa$ (non-trivial, i.e., embedding at least one conditional independence relation), there exists a decomposable model $\hat{\kappa}' = (K'_1 \bullet K'_2 \bullet \ldots \bullet K'_{k'})_\kappa$, such that there exists a pair of variables which are conditionally independent given the rest of variables in the model in the case of model $\hat{\kappa}$, but not in the case of model $\hat{\kappa}'$.*

For further clarification of notions, proofs and simple examples, see again [4].

2.3 Likelihood-ratio Statistics

For testing whether the compositional model $\hat{\kappa}$ sufficiently faithfully approximates the original data distribution κ (both with variables from $\mathbf{X}_K$), the likelihood-ratio test statistic is defined by the formula

$$G^2 = 2 \sum_{x_K \in \mathbf{X}_K} \kappa(x_K) \log \frac{\kappa(x_K)}{\hat{\kappa}(x_K)}$$

and, under certain conditions, has χ^2 distribution with the appropriate number of degrees of freedom (see above). In the case of a likelihood-ratio statistic, the sample large enough for the approximation of χ^2 distribution is usually considered when the sample is at least five times larger than the number of cells in a contingency table [1].

2.4 Decomposition of Likelihood-ratio Statistic

Now we shall take an advantage of the decomposability of models in order to decompose[3] the G^2 test statistics. This method employs the neighborhood structure of decomposable models as described in Theorems 1 and 2.

Using the properties of logarithm, we can take advantage of neighboring models and arrive at formula

$$G^2_{\hat{\kappa}'} = G^2_{\hat{\kappa}} + 2 \sum_{x \in \mathbf{X}_{K_i}} \kappa(x_{K_i}) \log \frac{\kappa(x_{K_i \setminus \{\ell,m\}})\kappa(x_{K_i})}{\kappa(x_{K_i \setminus \{\ell\}})\kappa(x_{K_i \setminus \{m\}})} \tag{1}$$

which allows the enumeration of the likelihood-ratio statistic using a pre-computed value of the neighboring model and employing only local computations.

2.5 Degrees of Freedom for Likelihood-Ratio Decomposition

The number of degrees of freedom for the likelihood-ratio statistics in the model $\hat{\kappa}$ are given by formula

$$df = \prod_{k \in K} r_k - 1 - \sum_{i=1}^{n} \left(\prod_{j \in K_i \setminus U_i} r_j - 1 \right) \cdot \prod_{j \in K_i \cap U_i} r_j$$

where symbol r_k denotes the number of categories for the corresponding variable. If we introduce a new conditional independence relation among the pair of variables with indices $\ell, m \in K_i$, the change in the number of degrees of freedom Δdf can be computed from the previous values as follows

$$\Delta df = (r_\ell - 1)(r_m - 1) \prod_{j \in K_i \setminus \{\ell,m\}} r_j. \tag{2}$$

This change in the number of degrees of freedom for neighboring models (after introduction of one additional conditional independence relation) is in agreement with the results obtained in the case of hierarchical log-linear models and Bayesian networks [1, 9].

2.6 Akaike Information Criterion

The test statistic itself does not contain information about the number of parameters used for the representation of the model. This information is, in the case of hypothesis testing, employed in the form of degrees of freedom. But in the 1970s, , the Akaike information criterion was formulated on the basis of the parsimony principle(see [2]). It can be expressed in a form using likelihood–ratio statistics

$$AIC_{G^2} = G^2 - 2 \cdot df$$

where G^2 is the likelihood–ratio and df is the number of degrees of freedom.

Now, in the case of neighboring decomposable compositional models, thanks to the above-expressed Formulae (1) and (2) we can again locally compute the value of Akaike information criterion and use it for search among models.

[3]The word decomposable (decompose) is used here in an ambiguous manner; it has two different meanings. The decomposability of a model is a structural property (characterizable by the RIP property). But the decomposability of a test statistic means the possibility to perform only local computations and hence take advantage of previous computations in the process of searching for a suitable model.

2.7 Suboptimal Search among Decomposable Models

The search algorithm uses a simple idea of limiting the number of decomposable models tested in each iteration to a certain limit k. The algorithm is based on the use of neighborhood theorems 1 and 2.

Suboptimal search using the test criterion algorithm starts with the saturated model (with no conditional independence relation introduced) and then the three steps described below follow.

1. *Generate all possible decomposable models with one additional conditional independence relation between a pair of variables.*
2. *Choose k models with the lowest values of the criterion.*
3. *Repeat steps 1 and 2 as long as the values of the criterion decrease.*

Obviously, we obtain only a suboptimal solution due to the greedy character of the algorithm, since it does not search the entire space of decomposable models.

3 Study Data 2011–2016

In the presented paper we analyze anonymized data from the university information system InSIS concerning 1827 Bachelor Degree students of a management study program on Faculty of Management in Jindřichův Hradec in the years 2011–2016. For the sake of clarity, we cover only results of obligatory subjects (without languages and sports). I.e., the analysis comprises the factors and obligatory study subjects summarized in Table 1.

Variable name	Meaning	Values
Gender	Gender of the student	female, male
State	Citizenship of the student	Czech Rep., Slovakia, Ukraine, other
Form	Study form	face-to-face, blended learning
HS type	Type of secondary school	Gymnasium, Business Academy, other
61APD0	Data Analysis and Presentation	1,2,3,4+,4,–,O
61EKA0	Business Economics	1,2,3,4+,4,–,O
61EKM0	Economics for Managers	1,2,3,4+,4,–,O
61FIM0	Finance for Managers	1,2,3,4+,4,–,O
61INM0	Informatics for Managers	1,2,3,4+,4,–,O
61MAN0	Management	1,2,3,4+,4,–,O
61MAT0	Mathematics for Managers	1,2,3,4+,4,–,O
61MKT0	Marketing	1,2,3,4+,4,–,O
61MLZ1	Human Resource Management 1	1,2,3,4+,4,–,O
61MND0	Managerial Skills	1,2,3,4+,4,–,O
61MOP0	Operations Management	1,2,3,4+,4,–,O
61MPR0	Project Management	1,2,3,4+,4,–,O
61OMP0	Operations Man., Processes and Supply Chains	1,2,3,4+,4,–,O
61PRA0	Law for Managers	1,2,3,4+,4,–,O
61SMD0	Study and Managerial Skills	1,2,3,4+,4,–,O
61SOC0	Sociology	1,2,3,4+,4,–,O
61UZD0	Accounting and Tax Bases	1,2,3,4+,4,–,O
61ZPC0	Basic PC skills (prerequisite test)	Z,NZ,–,O
61ZSM0	Basic Secondary School Math. (prereq. test)	Z,NZ,–,O
61BPMN	Bachelor's Thesis Defense	1,2,3,4,–,O
61SZBM	Major Field State Exam	1,2,3,4,–,O

Table 1 Obligatory subjects and other analyzed factors

3.1 Results of Structure Learning

The procedure of the structure learning was started with the full model (without any conditional independence relation). It continued with 236 steps of accepted model simplification, with the resulting model characterized by

$$G^2 = 44,569.5; \qquad df \doteq 8.8 \cdot 10^{17}; \qquad AIC = -13,609.9; \qquad \Delta_{AIC} = 4,485.1.$$

The resulting suboptimal solution consists of 21 marginal distributions given by the variables listed in Table 2

(Gender,Form,HStype,61SZBM)	(Gender,Form,61EKA0,61SZBM)	(Gender,Form,61FIM0,61ZPC0)
(Gender,Form,61MND0,61ZPC0)	(Gender,Form,61MPR0,61SZBM)	(Gender,Form,61SZBM,61UZD0)
(Gender,Form,61SZBM,61ZPC0)	(State,Form)	(Form,61SZBM,61ZSM1)
(Form,61BPMN,61SZBM)	(61APD0,61SZBM)	(61EKM0,61ZPC0)
(61INM0,61ZPC0)	(61MAN0,61ZPC0)	(61MAT0,61SZBM)
(61MKT0,61ZPC0)	(61MLZ1,61ZSM1)	(61MOP0,61SZBM)
(61PRA0,61ZPC0)	(61SMD0,61ZPC0)	(61SOC0,61ZPC0)

Table 2

The list of variables defines a set of low-dimensional distributions and can be depicted by a hypergraph for the sake of visualization. But since the hypergraph would contain 21 hyperedges in our case, the depiction could not be clearly arranged. Therefore, we decided to visualize the structure of the model using an (ordinary) undirected graph (see Figure 1). In such a graph, marginals are represented by cliques (maximal sets of vertices with edges connecting each other).

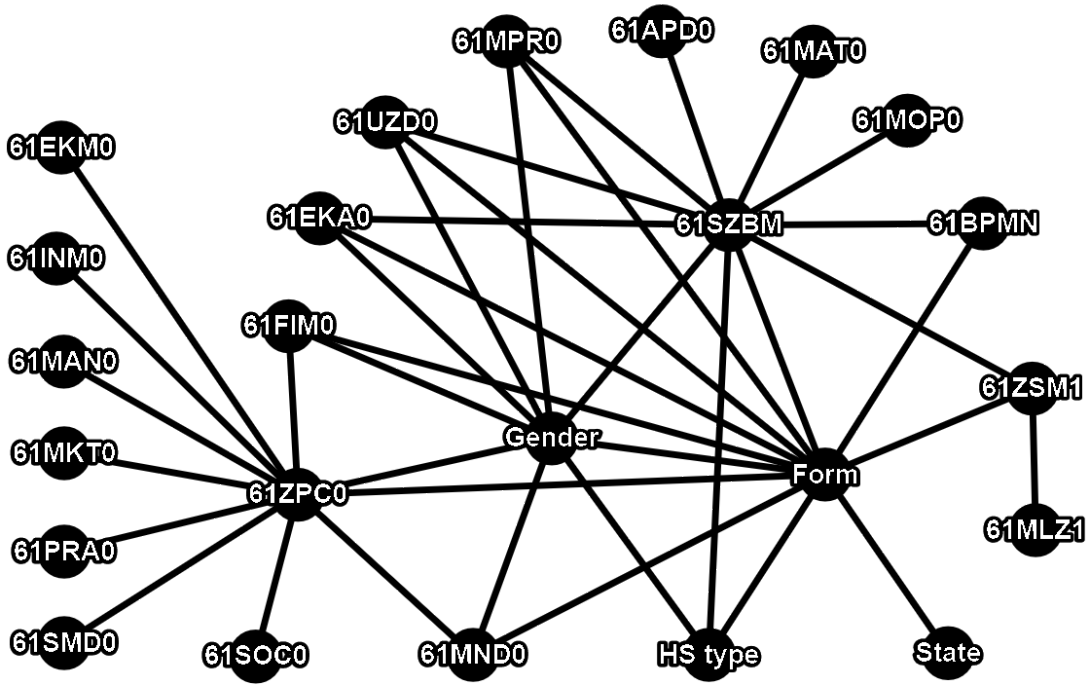


Figure 1 Graphical representation of the resulting compositional model.

3.2 Interpretation of Two Interesting Results

Within the limited space of the presented paper, let us show several interesting results. Table 3 represents one of the building stones of resulting compositional models, particularly, conditional distribution for the relation between the state and study form of students.

Form	Czech Rep.	Slovakia	Ukraine	Other
Blended learning	0.421	0.329	0.278	0.388
Face-to-face	0.579	0.671	0.722	0.612

Table 3 Dependence of chosen study form on state.

We can see that the blended learning form is more usual in the case of Czech students; the lowest frequency appeared in the case of students from Ukraine. The possible reason is that the cohort of students completing their

university studies at a higher age come nearly only from the Czech Republic. Other blended learning students are those who, at an age of about twenty, need to work because of insufficient financial resources of their families.

Let us show one other interesting difference for the students in face-to-face and blended learning forms. Table 4 shows different grades obtained in Bachelor's thesis defense and the major field state exam in both forms of study. As expected, face-to-face students appear to be slightly more successful and obtain better grades. The difference is more significant in the case of the state exam.

Face-to-face						Blended learning					
	61BPMN						61BPMN				
61SZBM	1	2	3	4	Sum	61SZBM	1	2	3	4	Sum
1	0.035	0.021	0.007	–	0.063	1	0.029	–	–	–	0.029
2	0.196	0.231	0.098	–	0.524	2	0.059	0.147	0.176	–	0.382
3	0.098	0.140	0.133	0.007	0.378	3	0.235	0.118	0.118	0.029	0.500
4	–	0.028	0.007	–	0.035	4	–	0.029	0.059	–	0.088
Sum	0.329	0.420	0.245	0.007	1.000	Sum	0.324	0.294	0.353	0.029	1.000

Table 4 Dependence of results in 'Bachelor's Thesis Defense' and 'Major Field State Exam' on study form.

4 Conclusions

We presented simplified study data of Bachelor Degree students from the Faculty of Management in Jindřichův Hradec using the apparatus of probabilistic compositional models. This approach makes use of significant computational advantages of decomposable compositional models, decomposable test criterion and performing of local computations in the algorithm searching within the space of possible compositional models. A significant simplification of the data file and reduction of the set of factors affecting the resulting model rank among the drawbacks of the presented results. Moreover, the presented algorithm is a basic one and suffers from both its greedy character and its start from the full (saturated) model. As such, the algorithm can be employed as a starting one and the results can be made more accurate taking advantage of both neighborhood theorems in a more sophisticated (e.g., oscillating) approach.

Besides the resulting model, we have shown several interesting observations following from the structure of the data. But it is apparent that the space dedicated to this conference paper does not provide the possibility to work further with the resulting model and to present the operations of marginalization, conditioning and intervention. Theoretical framework for such operations was published in Bína and Jiroušek [6].

Acknowledgements

The research was supported by GA ČR under Grant No. 15-00215S.

References

[1] Agresti, A.: *Categorial Data Analysis*. John Wiley & Sons Inc., New York, 2002.

[2] Akaike, H.: A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**(6) (1974), 716-723.

[3] Bína, V.: Greedy Search among Decomposable Models. In: *Proceedings of The Fourteenth IASTED International Conference on Artificial Intelligence and Soft Computing* (I. Andreadis, M. Zervakis, eds.)., Greece, 2011, pp. 280–287.

[4] Bína, V.: *Multidimensional Probability Distributions: Structure and Learning: Disertation Thesis*. Fakulty of Management, VŠE Praha, 2011.

[5] Bína, V. and Jiroušek, R.: Marginalization in Multidimensional Compositional Models, *Kybernetika* **42**(4) (2006), 405-422.

[6] Bína, V. and Jiroušek, R.: On Computations with Causal Compositional Models, *Kybernetika* **51**(3) (2015), 525-539.

[7] Jiroušek, R.: Foundations of Compositional Model Theory, *International Journal of General Systems* **40**(6) (2011), 623–678.

[8] Koller, D. and Friedman, N.: *Probabilistic Graphical Models: Principles and Techniques*. Massachusetts Institute of Technology, 2009.

[9] Neapolitan, R. E.: *Learning Bayesian Networks*. Prentice Hall, 2003.

Selection of the suitable building savings in the Czech Republic using multicriteria evaluation method

Adam Borovička[1]

Abstract. Building savings is one of the most favorite saving products in the Czech Republic. The purpose of that product is ongoing savings of free financial resources for future consumption, eventually for getting a loan. Building savings is offered by many banks in our country. Apart from a given state contribution, other characteristics (deposit or loan interest rate, account management fee, contract fee etc.) can be different among the banks. All relevant criteria should be taken into account in the process of the building savings selection in order to make a responsible and judicious decision. To choose the most suitable building savings, I propose to apply the quantitative approach of multicriteria evaluation of alternatives. For this purpose, enhancing modifications and combinations of ELECTRE I and ELECTRE III methods are introduced. In the practical part of the paper, two savings strategies are specified (with/without purpose of getting a loan). For each strategy, the most suitable alternative of building savings is selected from the group of building savings of ten (savings) banks in the Czech Republic by the introduced method. The different results are analyzed discussed.

Keywords: building savings, ELECTRE methods, multicriteria evaluation.

JEL Classification: C44, G11
AMS Classification: 90B50

1 Introduction

Many people in the Czech Republic responsibly think of their future financial needs. These needs could be invoked by important investment (e.g. in housing, car), more usual consumption (electronics, clothes etc.) or financial old-age security. For these purposes, free financial resources must be continuously saved. One way of savings money is through a well-known product building savings. It is a type of savings which is contributed by the state. Namely therefore this product is very popular in the Czech Republic. Moreover, the survey of the agency International Business & Research Services for Association of Czech Building Savings Banks [15] confirms the ever-increasing demand for this product in the Czech Republic.

In the Czech Republic, ten (savings) banks offer the building savings product. These products can have various parameters. Then the fundamental question is which product is the best for the client. Selection of the most suitable building savings is the main aim of this article. In practice, many people arrange the building savings without more detailed analysis by the bank where they have their current account. This way is easy and too little time-consuming. The main disadvantage of this approach is that some better alternatives can be omitted. Other often negative aspect of the building savings selection is that the clients are not usually interested in all important characteristics of this product. They usually see only attractive state contribution and there's an end of it! To eliminate these abuses, I propose the complex quantitative multicriteria analysis of all available building savings in the Czech market.

To fulfill the aim, the multicriteria evaluation method is introduced. The algorithm of this methods is significantly affected by the well-known ELECTRE I and ELECTRE III methods. Modifications and combinations of these methods assist to make a reasonable decision in this concrete situation. Used method must be able to select "the best" alternative (eventually to divide in "good" and "bad" alternatives), or to make a ranking of alternatives on the basis of all selected characteristics. Because the most clients are outsiders, the method must be comprehensible for a wider range of users. The practical part of this article tries to represent as many real cases as possible. Therefore two typical cases are specified. One of them is the client whose primary aim is the yield from the deposit. The second typical client plans to get a (smaller) loan from the building savings in the future. It is expectable that these different preferences can affect a selection. The proposed multicriteria method satisfying all aforementioned requirements is applied to both strategies. The different results are analyzed and disccused.

The structure of this article is as follows. Introduction is followed by the section containing a treatise of multicriteria evaluation method that is proposed for building savings selection (not only) in the Czech Republic. This

[1] University of Economics, Department of Econometrics, W. Churchill Sq. 4, Prague, Czech Republic, adam.borovicka@vse.cz

real problem is described in Section 3. The evaluative criteria are specified. Two types of client and their preferences are determined. After data collection, the suitable buildings savings are selected. In the end, the article is summarized and some stimulations for future research are outlined.

2 Multicriteria evaluation method

In a few last decades, many multicriteria evaluation methods have been developed. These methods can be classified by the principle of evaluation of alternatives into 3 groups. One of them is created by methods using the utility function (WSA, AHP, ANP) [see more e.g. 4]. Another large group contains the multicriteria evaluation methods based on the concept of minimization of a distance from the ideal alternative. The best-known method is apparently TOPIS [5]. These methods differentiate in the applied metric. The third good-sized group is represented by methods using a concept of the preference scoring relation. The well-known are AGREPREF method [7], group of methods ELECTRE (e.g. ELECTRE I [13] and III [14]), or PROMETHEE (e.g. PROMETHEE I and II [1]). Besides the "basic" mentioned methods, another multicriteria evaluation methods are known. Of course, some methods are variously modified, improved or combined (in normalization technique, importance of criteria, fuzzification or stochastic input data) mostly due to solve a real decision making problem satisfactorily. Analogous way is also chosen in this article.

Each method has usually some advantages and disadvantages which are related to the form of algorithm and its applicability in the real decision making situation. Firstly, the construction of the utility function can be difficult for many users because the utility can be too abstract and impalpable concept for them. On the other side, a determination of the utility is sometimes simplified by the integrated formulae in the algorithms. Secondly, the distance measurement concept is mostly better comprehensible. But the result can be very vulnerable to the used technique of distance calculation. Thirdly, many methods based on the preference scoring relation require information from a decision maker in the form of the threshold value(s) which can be problematic namely for less experienced users. From the perspective of my practical problem, the third group of methods is preffered. The first two group of methods usually provide a ranking of alternatives. But it is not the main aim of the specified problem. The client primarily wants to know "the best" alternative, eventually "good" and "bad" alternatives. The full ranking can be a supporting information. However, this is rather only slight benefit of methods from the third group. More essential fact is that many methods from the first two groups (e.g. WSA or TOPSIS) use the normalization technique of the criteria values which distort the original data (especially relative relationships among them).

After research of the multicriteria evaluation methods, I decided to apply the method based on the preference scoring relation. Inspiration was namely in ELECTRE I and ELECTRE III methods. Some parts of the algorithms are taken over, some are improved and combined. The result is a method that can divide the alternatives in "good" and "bad", to select "the best" one, as well as to make their ranking. Any additional information (threshold values, specification of the utility function) from the users is not demanded. Normalization technique does not distort the input data. Another advantage is that the algorithm of this method is user-friendly. Its implementation is quite easy. In this article, the emphasis is on a solving of real decision making problem, so the algorithm of multicriteria evaluation method is described only briefly in terms of a few following steps.

Step 1: Let $\mathbf{Y} = (y_{ij})$ be the matrix with the elements $y_{ij} (i = 1, 2, ..., n;\ j = 1, 2, ..., k)$ representing the evaluation of the *i*-th alternative by the j-th criterion. Importance of the j-th criterion is quantified as a weight w_j. As in ELECTRE III, the following couple of sets of the criteria indices for each couple of alternatives is specified

$$I_{iPj} = \left\{ r \vee s \mid y_{ir} > y_{jr},\ y_{is} < y_{js};\ r \in I^{\max}, s \in I^{\min} \right\} \qquad i, j = 1, 2, ..., n;\ i \neq j$$
$$I_{jPi} = \left\{ r \vee s \mid y_{jr} > y_{ir},\ y_{js} < y_{is};\ r \in I^{\max}, s \in I^{\min} \right\} \qquad i, j = 1, 2, ..., n;\ i \neq j \quad ,$$

where set $I^{\max}$, or $I^{\min}$ contains the indices of maximizing, or minimizing criteria.

Step 2: Two matrix of the grades of preference are determined. The matrix $\mathbf{S} = (s_{ij})$ is formulated as in the ELECTRE III method. The matrix **R** is inspired by the ELECTRE I concept. The main distinction and also advantage is that the formula works with the criteria importance and normalized criteria values. Such an indicator has a better predictive value. Its element r_{ij} is calculated for each couple of alternatives *i* and *j* ($i, j = 1, 2, ..., n$) as follows

$$r_{ij} = \frac{\sum_{h \in I_{iPj}} \left(w_h \mid y'_{ih} - y'_{jh} \mid \right)}{\sum_{h=1}^{k} w_h \mid y'_{ih} - y'_{jh} \mid} \qquad I_{iPj} \neq \emptyset,$$

$$, r_{ij} = - \qquad i = j,$$

$$r_{ij} = 0 \qquad \text{else}$$

where y'_{ih}, or y'_{jh} $(i, j = 1, 2, ..., n; h = 1, 2, ..., k)$ is a normalized criteria value.

Step 3: Aggregate preference of the i-th alternative over the j-th alternative is determined by the following rule

$$s_{ij} > s_{ji} \wedge r_{ij} > r_{ji} .$$

This rule is a modification of ELECTRE III approach. The threshold values are eliminated. Then effective ("good") alternative is such an alternative that has the highest discrepancy between number of alternatives over which is preferred and number of alternatives that are preferred over it. Other alternatives are *ineffective* ("bad"). This concept is a combination of ELECTRE I and III approaches. This improvement namely eliminates a crucial drawback of ELECTRE I concept that the effective alternative may not exist. If there are more effective alternatives then they can be distinguished by the additional procedure in order to select "the best" one. The full ranking of alternatives can be made by the similar procedure.

3 Building savings selection

In this section, two typical situations with the building savings are described. At first, let me determine the common conditions for both strategies. The early amount of saved money is mostly influenced by the amount of state contribution. Its maximum level 2 000 CZK is for a deposit 20 000 CZK. This most frequent alternative is considered in this analysis. The building savings is typically for 6 years. The expected final saved amount after 6-year period is about 135 000 CZK. Then, the target amount is standardly set with some reserve to 150 000 CZK.

In order to cover as many real cases as possible, two types of clients are specified - *client focused on the yield* and *client focused on getting a loan*. The main aim of the first client is to save free financial resources for future usage (for housing, electronics, usual consumptions as meals, clothes etc.). The main aim of the second client is to save money for getting a loan that is usually preferable than "classic" loans from the banks. This loan from the building savings is often used for a part of the price of a real property which cannot be covered by a mortgage. Nowadays, new conditions in the mortgage supply in the Czech Republic do not enable 100% mortgage. So this type of loan will have greater and greater meaning in our country. In this analysis, the amount of a loan is 500 000 CZK which can be approximately 10 % of real property price which cannot be covered by the mortgage.

In the Czech Republic, there are over 20 (savings) banks providing personal finance services. Ten (savings) banks offer in their products portfolio the building savings: *Českomoravská stavební spořitelna* (ČMSS), *Československá obchodní banka* (ČSOB), *Komerční banka* (KB), *Modrá pyramida stavební spořitelna* (MPSS), *Moneta Money Bank* (MMB), *Oberbank* (OB) *Poštovní spořitelna* (PS), *Raiffeisen stavební spořitelna* (RSS), *Stavební spořitelna České spořitelny* (SSČS) and *Wüstenrot – stavební spořitelna* (WSS). Abbreviation in the brackets will be used in the next sections to save space and preserve clarity.

The main aim of this section is to select a suitable building savings for each strategy. For this purpose, the introduced concept of multicriteria evaluation method will be applied.

3.1 Criteria

To make a complex analysis and responsible decision, the alternatives of building savings should be evaluated from more perspectives. Because the analysis will be full, 6 representative characteristics of the building savings are presented. The *contract fee* (CF) is an input cost for a signing of the contract. It is usually calculated as a percentage of the target amount. The *account management fee* (ACF) is a yearly cost connected with an account management. The *deposit interest rate* (DIR) is a yearly interest rate from deposits of the clients. The *loan interest rate* (LIR) is an early interest rate from the pertinent loan. The *loan contract fee* (LCF) is an input cost for getting a loan from the building savings. The *loan account management fee* (LAMF) is a yearly cost connected with a management of the loan account. It is obvious that the first three criteria are directly related to the actual building savings. Other three criteria are related to a potential loan from the building savings. Of course, maybe the most

interesting characteristics of the building savings seems to be omitted. But the *state contribution* is the same for all building savings in these two described real cases.

3.2 Preferences of the clients and appropriate data

Let me start with the most typical client who is primarily focused on the yield. This client regularly saves money for six years in order to gain a maximum profit. It is no wonder that besides state contribution the most important characteristics is a deposit interest rate. Other characteristics decreasing a future yield from the building savings are contract fee and account management fee. These factors are also important for this client, but naturally not so much compared with the interest from deposits. Indeed, these interests form a considerable part of the total revenues. This type of client has no plans getting a loan so the characteristics regarding a loan are not important. If by chance s/he will need a loan after all than, the most important characteristics connected with a loan will be logically the loan interest rate.

Weights of the criteria are calculated by the scoring method [4] on the basis of the described client's preferences. The client assigns an integer score from interval $\langle 0,10 \rangle$, where 0 represents the lowest importance and 10 the highest importance of the criterion. The scores and weights are shown in the following table (Table 1).

Criterion	**CF**	**AMF**	**DIR**	**LIR**	**LCF**	**LAMF**
Score	4	3	10	2	1	1
Weight	0.190	0.143	0.476	0.095	0.048	0.048

Table 1 Scores and weights of the criteria for the client focused on the yield

For each type of client, the data are different. For the client focused on the yield, the data is in the following form (Table 2). Used abbreviations are specified in section 3.1. All data is collected from the web pages of particular (savings) banks [2,3,6,8,9,10,11,12,16,17].

Bank	**CF [CZK]**	**AMF [CZK]**	**DIR [%]**	**LIR [%]**	**LCF [CZK]**	**LAMF [CZK]**
ČMSS	1500	330	0.50	3.50	5000	330
ČSOB	1500	330	0.50	3.50	5000	330
KB	1500	300	1	3.49	4400	300
MPSS	1500	300	1	3.49	4400	300
MMB	1500	324	1	3.99	0	0
OB	1500	324	1	3.99	0	0
PS	1500	330	1	4.30	5000	330
RSS	1500	360	1	x	x	x
SSČS	495	325	1	2.99	0	325
WSS	1500	324	1	3.99	0	0

Table 2 Data of the buildings savings for the client focused on the yield

Because differences of the criteria values are not so high, some building savings can actually have the same parameters. There are three cases. The first one is ČMSS and ČSOB, the second is KB and MPSS and the third MMB, OB and WSS. We can see some connections in the particular cases. KB is a main shareholder of MPSS. KB actually offers the building savings through MPSS. OB and WSS are partnership banks. Therefore, they offer the same savings product. The contract fee is 1 % from the target amount (1500 CZK in our case). However, SSČS offer online contract making for a lower price 495 CZK. Deposit interest rate is usually 1 % from deposits. ČSOB and ČMSS have just only 0.5 %. RSS and WSS also offer a lower deposit interest rate that engage a lower loan interest rate. Because a loan interest rate is not so important factor for this saving strategy, a deposit interest rate is set to the maximum level. Building savings from RSS has 1 % deposit interest rate only in the case when a loan is not available. Unavailability of the loan is not a problem for this type of client. Then for the multicriteria analysis, the value of „loan" characteristics are set to the inconvenient levels to be dominated by other alternatives in these characteristics. Loan contract fee is free for MMB, OB, SSČS and WSS. This fee is set to 4400 CZK by KB and MPSS. ČMSS, ČSOB and PS have a loan contract fee as 1 % from the amount of a loan (5000 CZK in our case).

The second very often case is the client who assumes that a loan will be needed in the future. It is obvious that the most important characteristics is a loan interest rate. Interest is clearly the greatest cost of the loan. Other cost is loan contract fee and loan account management fee that are not just so important because these costs are not so

high in comparison with the interest from a loan. Deposit interest rate is not so important criterion compared with the previous type of client. But it is not possible to say that it is insignificant viewpoint in the analysis. Interest of the deposits is also significant, but naturally not so much as interest from a loan. Moreover, the building savings of some banks offer a lower loan interest rate under condition of a lower deposit interest rate. The purpose of this building savings causes that the cost connected with the contract and account management is relatively insignificant. It is expectable that these costs will be somewhat less important than the "administrative" cost regarding a loan.

Weights of the criteria are also calculated by the scoring method on the basis of the client's preferences that are (as in the previous case) determined on the basis of my considered opinion and a survey among (potential) clients in my immediate neighborhood. The scores and weights are presented in this table (Table 3).

Criterion	**CF**	**AMF**	**DIR**	**LIR**	**LCF**	**LAMF**
Score	3	2	5	10	4	3
Weight	0.111	0.074	0.185	0.371	0.148	0.111

Table 3 Scores and weights of criteria for the client focused on getting a loan

The data differs from the previous client in 3 products – building savings OB, RSS and WSS. Other values are the same. Then only three mentioned cases are presented in the following table (Table 4).

Bank	**CF [CZK]**	**AMF [CZK]**	**DIR [%]**	**LIR [%]**	**LCF [CZK]**	**LAMF [CZK]**
OB	1500	324	0.50	1.99	0	0
RSS	1500	360	0.50	3.50	0	360
WSS	1500	324	0.50	1.99	0	0

Table 4 Data of the buildings savings for the client focused on getting a loan

As mentioned earlier, OB and WSS offer a lower deposit interest rate with a lower interest rate for a loan. Then a deposit interest rate 0.5 % enables to reach 1.99 % loan interest rate. RSS also offers 0.5 % deposit interest rate, but a loan interest rate is significantly higher at the value 3.50 %. Now we have 3 couples of building savings with the same parameters. One couple is ČMSS and ČSOB, the second is KB and MPSS, and the third one is created by OB and WSS.

3.3 Suitable building savings for both strategies

Primarily, the non-dominancy test is made. The dominated alternatives are building savings from ČMSS, ČSOB, PS and RSS for both strategies. These alternatives are not attractive for the client. To analyze all alternatives of building savings, the introduced multicriteria evaluation method is applied namely through MS Excel and its add-in for multicriteria analysis Sanna.

The "best" (effective) alternative for the client focused on the yield is the building savings from SSČS. This result was expectable and is caused by the following facts. This building savings has the greatest deposit interest rate which is the most important criterion. However, this value is embodied by many other building savings. Second reason is that this building savings has the lowest contract fee which is the second most important criterion. This value is significant priority to the others. The result is also supported by a very good level of the loan interest rate and loan contract fee. If the client wants to have a complex image about the alternatives evaluation, additional procedure can be applied to make a full ranking of building savings. The ranking is as follows: 1. *SSČS*, 2.-3. *KB*, *MPSS*, 4.-5. *MMB*, *OB*, 6. *WSS*, 7. *PS*, 8. *RSS*, 9.-10. *ČMSS*, *ČSOB*. The second, or third place of building savings of KB and MPSS is namely caused by their best value of an account management fee that is the third most important characteristics. Of course, four dominated alternatives are on the last four places. PS is in face of RSS due to a lower account management fee. The worst position is assigned to ČMSS and ČSOB due to the lowest deposite interest rate.

As mentioned above, the data is changed a little for the client focused on getting a loan. It is obvious from the data list that the main question is whether the building savings SSČS will be again the "best" alternative or the building savings of WSS, or OB with the best value of all three "loan" characteristics wins. For their win, the importance of a deposit interest rate must significantly decrease and the importance of a loan interest rate must significantly increase. It is important to be the weight of a deposit interest rate lower than the weight of a loan interest rate. In this situation, the building savings WSS, or OB can win. These assumptions are adequately met so the "best" (effective) alternatives are just these building savings. Namely to compare with the result for the client focused on the yield, the full ranking is: 1.-2. *WSS*, *OB*, 3. SS*ČS*, 4. *MMB*, 5.-6. *KB*, *MPSS*, 7. *RSS*, 8.-10. *ČMSS*,

ČSOB, PS. The building savings of MMB gets ahead of the building savings from KB and MPSS. Despite a greater loan interest rate of building savings MMB, the building savings MMB is primarily better due to the best value of loan contract fee and loan account management fee. The building savings from PS drops in the order due to very bad value of all three "loan" characteristics. This building savings is not identical with the building savings of ČMSS and ČSOB, however all saving alternatives share the last place. The main reason is that these alternatives accordingly have no alternatives before which would be preferred.

4 Conclusion

The main mission of this article is to select the suitable building savings in the Czech Republic where this savings product is very popular. For this purpose, multicriteria evaluation method is introduced. This method enables to make a complex analysis of the building savings reflecting all important perspectives. In the practical part, two typical situations with the building savings are formulated (strategy focused on the yield and strategy focused on getting a loan.). The results show that the client's preferences play an important role in the process of building savings selection. The most suitable product is different in both strategies. In the future research, some special cases will be analyzed. For instance, some (savings) banks offer a higher deposit interest rate under particular conditions. Further, some (savings) banks make a contract for free for children, people over 55 years old etc. These and other aspects can affect the result. For all planned analyzes, detailed survey will be carried out using a ques-tionnaire focusing on expressing preferences (potential) clients.

Acknowledgements

The research project was supported by Grant No. IGA F4/57/2017 of the Internal Grant Agency, Faculty of Infor-matics and Statistics, University of Economics, Prague.

References

[1] Brans, J. P., Vincke, P.: A preference ranking organization method. *Management Sciences* **8** (1985), 647–656.
[2] ČMSS – stavební spoření. Retrieved April 1, 2017, from https://www.cmss.cz/stavebni-sporeni.
[3] ČSOB – stavební spoření. Retrieved April 1, 2017, from https://www.csob.cz/portal/lide/produkty/sporici-produkty/stavebni-sporeni.
[4] Fiala, P.: *Modely a metody rozhodování*. Oeconomica, Praha, 2013.
[5] Hwang, C. L., Yoon, K.: *Multiple Attribute Decision Making – Methods and Applications, A State-of-the-Art Survey*. Springer-Verlag, New York, 1981.
[6] Komerční banka – stavební spoření. Retrieved April 1, 2017, from https://www.kb.cz/cs/sporeni-a-investovani/sporeni/stavebni-sporeni/.
[7] Lagrèze, E. J.: La modélisation des préférences: Préordres, quasiordres et relations floues. *Thèse de doctorat*, Université de Paris V, 1975.
[8] Modrá pyramida – stavební spoření. Retrieved April 1, 2017, from https://www.modrapyramida.cz/online-nas-troje/stavebni-sporeni.
[9] Moneta Money Bank – stavební spoření. Retrieved April 1, 2017, from https://www.moneta.cz/lide/sporeni/stavebni-sporeni.
[10] Oberbank – stavební spoření. Retrieved April 1, 2017, from http://www.oberbank.cz/OBK_webp/OBK/ober-bank_cz/Privatni_klienti/Sporeni_a_investovani/Stavebni_sporeni/index.jsp.
[11] Poštovní spořitelna – stavební spoření. Retrieved April 1, 2017, from https://www.erasvet.cz/fyzicke-osoby/sporeni-a-investice/stranky/stavebni-sporeni/popis-produktu.aspx.
[12] Raiffeisen Bank – stavební spoření. Retrieved April 1, 2017, from https://www.rb.cz/osobni/sporeni-a-investovani/sporeni/stavebni-sporeni.
[13] Roy, B.: Classement et choix en présence de points de vue multiples (la méthode ELECTRE). *La Revue d'Informatique et de Recherche Opérationelle (RIRO)* **8** (1968), 57–75.
[14] Roy, B.: ELECTRE III: algorithme de classement basé sur une représentation floue des préférences en présence des critères multiples. *Cahiers du CERO* **20** (1978), 3–24.
[15] Sovová, E.: Nový trend: Češi mění zaběhané postoje, kam ukládat peníze. Retrieved March 13, 2017, from http://fi-nance.idnes.cz/barometr-sporici-produkty-rust-a-pokles-obliby-f5j-/sporeni.aspx?c=A170310_103941_sporeni_sov.
[16] Stavební spořitelna České spořitelny – stavební spoření. Retrieved April 1, 2017, from https://www.burinka.cz/cs/produkty/stavebni-sporeni/.
[17] Wüstenrot – stavební spoření. Retrieved April 1, 2017, from https://www.wuestenrot.cz/sporeni/prosporeni.

A stochastic-integer programming approach to tactical fixed interval scheduling problems

Martin Branda [1]

Abstract. Fixed interval scheduling (FIS) problems arise in many areas of economics and industry where jobs with processing intervals known in advance are assigned to available machines with forbidden preemption. A special case of such problems appears in the personnel task scheduling where the decision maker (manager) is looking for a minimal number of workers to cover all prescribed shifts. This problem can be classified as a tactical fixed interval scheduling. In this paper, we focus on scheduling of jobs with uncertain processing intervals where the finishing times are modelled as random variables with a known probability distribution. We provide a stochastic integer programming formulation with a joint chance constraint which ensures the reliability of the resulting schedule. We propose an iterative decomposition algorithm for a reformulation where the partial operational FIS problems can be solved as the min-cost network flow problems. The performance of the algorithm is verified on simulated instances.

Keywords: Tactical fixed interval scheduling, random processing intervals, stochastic programming, integer programming, decomposition algorithm.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Stochastic fixed interval scheduling (FIS) problems investigated in this paper can be characterized as follows. The number of jobs together with their starting and prescribed finishing times are known in advance. However, the finishing times are uncertain and the uncertainty is represented by stochastic delays. These delays can incorporate unpredictable complications during processing the jobs or machine failures and necessary reparations or maintenance. We assume that each machine can process at most one job at any time and each job is processed by exactly one machine. Moreover, if a job is assigned to a particular machine, it cannot be transferred to another machine, i.e. preemption is not allowed. Thus, each machine operates independently of other machines. The operational FIS problem addresses assignment of jobs to available machines to maximize the schedule reliability, i.e. probability that it is feasible with respect to the realizations of the random delays. The tactical FIS problem, which will be our main concern, is to minimize the number of machines which are necessary to process the jobs under a prescribed level of reliability. Note that the FIS problems are related to personnel task scheduling problems, cf. Smet et al. [14]. Figure 1 shows two feasible assignments of 6 jobs to 3 machines.

Deterministic tactical fixed interval scheduling problems were considered by several papers. Kroon et al. [12] introduced a tactical variant of the FIS problem where the goal is to find the minimal number of machines to process all known jobs. They proposed a lower bounding procedure based on the Lagrangian duality and an upper bounding procedure relying on a greedy heuristic. There are two survey papers that summarize older results on various deterministic FIS problems, cf. Kolen et al. [10], and Kovalyov et al. [11]. Eliiyi [9] considered FIS problems which combine operational and tactical decisions, discussed the complexity of their special cases and proposed exact and heuristic algorithms. Zhou et al. [17] generalized the tactical problem by including the spread-time constraints which restrict the time for which a machine is available since the time when the first job was processed on it. They proposed an effective branch-and-price algorithm. Ng et al. [13] introduced an exact and several heuristic algorithms based on graph representations of the FIS problems.

Concerning the FIS problems under stochastic delays, we can refer to the following papers. Branda et al. [7] introduced an extended robust coloring formulation of the operational problem and showed that small and middle sized problems can be solved to optimality using a standard mixed-integer solver. For larger instances, they implemented a special tabu search algorithm. Branda and Hájek [4] dealt with the operational problems with heterogeneous machines where the distribution of delay can depend on the processing machine. The authors

[1]Charles University
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 83, Prague 186 75, Czech Republic
tel.: +420 221 913 404, fax: +420 222 323 316, branda@karlin.mff.cuni.cz

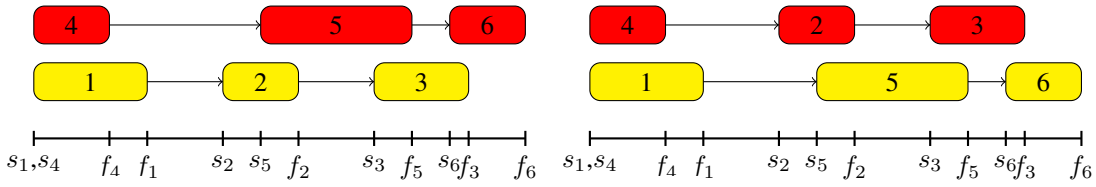


Figure 1 FIS schedules: two feasible assignments of 6 jobs to 2 machines

proposed network flow reformulations with side constraints which were shown helpful for solving larger instances to optimality.

The stochastic tactical FIS problem considered in this paper can be characterized as follows. All machines are identical, i.e. the processing times including the distribution of random delays do not depend on the selected machine. The probabilistic distribution of random delays is known precisely or at least its good approximation is available. The goal is to find the minimal number of (active) machines such that all jobs can be processed under a prescribed schedule reliability. We will employ an approach based on the min-cost network flow reformulations of the operational FIS problem which enables to obtain the best attainable reliability for a given number of machines and jobs quickly, see Branda [3]. This can be used by a decision maker to select the minimal number of machines with the prescribed schedule reliability by solving a sequence of the network flow problems. Good performance of such approach is verified in the numerical study.

The paper is organized as follows. A stochastic programming formulation of the tactical FIS problem is provided in Section 2. In Section 3, we review the min-cost network flow reformulation for the operational problem under independent delays which is crucial for solving the tactical problem. An extensive numerical study is provided in Section 4 where we summarize results obtained for 3000 simulated FIS instances. Section 5 concludes the papers with a summary of the obtained results and an outline of the future research directions.

2 Tactical fixed-interval scheduling problem

In this section, we provide a stochastic integer programming formulation of the tactical FIS problem. In particular, we will formulate a chance constrained problem (CCP) where a necessary schedule reliability is prescribed. The chance constrained problems are highly demanding in general and special purpose algorithms are necessary to solved them, see, e.g., Adam and Branda [1], Shapiro et al. [15].

Denote by s_j and $f_j(\xi)$ the starting and the uncertain (stochastic) finishing time for a job $j \in \mathcal{J} := \{1, \dots, J\}$, where ξ is used to denote the elementary events of a probability space $(\Omega, \mathcal{F}, P)$. We assume that the true finishing time is a sum of the prescribed finishing time f_j and a random delay $D_j(\xi)$. We denote by $\mathcal{T} = \{s_1, \dots, s_J\}$ the set of all starting times when it is sufficient to verify how many jobs are assigned to the machines. The set of available identical machines is denoted by $\mathcal{M} = \{1, \dots, M\}$. We assume that the number of available machines M is sufficient such that the problem is feasible. A prescribed level $\varepsilon \in (0, 1)$ corresponds to the allowed probability of violating the schedule feasibility. Binary decision variables $x_{jm} \in \{0, 1\}$ are used to assign the jobs to the machines, binary variables z_m identify the active machines. The tactical FIS problem is:

$$
\begin{aligned}
\min_{x_{jm}, z_m} \ & \sum_{m \in \mathcal{M}} z_m \\
\text{s.t. } & x_{jm} \le z_m, \quad j \in \mathcal{J}, \quad m \in \mathcal{M}, \\
& P\left(\xi \in \Xi : \sum_{j:\, s_j \le t < f_j(\xi)} x_{jm} \le 1,\ t \in \mathcal{T},\ m \in \mathcal{M} \right) \ge 1 - \varepsilon, \\
& \sum_{j:\, s_j \le t < f_j} x_{jm} \le 1, \quad m \in \mathcal{M}, \quad t \in \mathcal{T}, \\
& \sum_{m \in \mathcal{M}} x_{jm} = 1, \quad j \in \mathcal{J}, \\
& x_{jm} \in \{0, 1\}, \quad m \in \mathcal{M}, \quad j \in \mathcal{J}, \\
& z_m \in \{0, 1\}, \quad m \in \mathcal{M}.
\end{aligned}
\tag{1}
$$

The objective function represents the number of active machines. A straightforward generalization is possible by adding prices for using the machines. Then the objective corresponds to the expenses. The first constraints ensure that if a job is scheduled to a machine, then this machine is marked as active. The chance constraint places a minimal level of the schedule reliability. The next constraints guarantee that at most one job is assigned to a machine at any starting time and that any job is assigned to exactly one machine.

The solution strategy can be based on an observation made by Branda [3]: If we increase the number of available identical machines in the operational problem, then the highest attainable reliability cannot decrease. Thus, we can use a series of the operational FIS problems to obtain the highest reliability under a given number of machines and then to find the minimal number of machines necessary to reach the goal reliability $1-\varepsilon$ in the tactical problem. As we will review in the following section, the operational problem on identical machines can be solved quickly (in polynomial time) using a network flow reformulation.

3 Network flow reformulation for the operational problem

In this section, we focus on the following operational problem and its network flow reformulation.

$$\begin{aligned}
&\max_{x_{jm}} P\left(\xi \in \Xi : \sum_{j:\, s_j \le t < f_j(\xi)} x_{jm} \le 1,\ t \in \mathcal{T},\ m \in \mathcal{M}\right)\\
&\sum_{j:\, s_j \le t < f_j} x_{jm} \le 1, \quad m \in \mathcal{M}, \quad t \in \mathcal{T},\\
&\sum_{m \in \mathcal{M}} x_{jm} = 1, \quad j \in \mathcal{J},\\
&x_{jm} \in \{0,\, 1\}, \quad m \in \mathcal{M}, \quad j \in \mathcal{J}.
\end{aligned} \tag{2}$$

Compared with the tactical problem (1), the schedule reliability was transfered to the objective function. The constraints have the same meaning as in the tactical problem. To solve the tactical problem, we increase the number of available machines M by one in each iteration and solve the operational problem (2) until we reach the optimal probability at least $1-\varepsilon$. The corresponding number of machines M is then the optimal value for the tactical FIS problem. The proposed procedure is always finite because the necessary number of machines is bounded by the number of jobs J when the schedule has reliability 1 (each job can be assigned to different machine). We will show below that the operational problems (2) can be solved efficiently.

The min-cost network flow reformulation of the operational problem (2) was proposed by Branda [3]. The network (V, E) can be described as follows. There are two nodes for each job corresponding to the interval start $s_j \in V$ with demand $d_{s_j} = -1$ and prescribed end $f_j \in V$ with demand $d_{f_j} = 1$. These nodes are connected by an edge $(s_j, f_j) \in E$. There are two artificial nodes corresponding to the source $0 \in V$ with demand $d_0 = M$ and sink $2J+1 \in V$ with demand $d_{2J+1} = -M$. The source is connected by an edge with all starting nodes $(0, s_j) \in E$, whereas there is an edge from each finishing node to the sink node $(f_j, 2J+1) \in E$. Maybe the most important edges connect a finishing time with a starting time that is higher or equal, $\{f_i, s_j\} \in E$ if $f_i \le s_j$, i.e. these edges correspond to the jobs which can be assigned to the same machine. We denote these edges by $\overline{E} \subset E$. We assume that the delays are independent. Then the costs assigned to these edges are equal to

$$c_{f_j,s_i} = -\ln(P(D_j(\xi) \ge s_i - f_j)).$$

Costs for all other edges are set to zero. The corresponding min-cost network flow problem can be then formulated as follows where the reliability in (2) can be obtained by applying $\exp\{-\cdot\}$ to the optimal value of:

$$\begin{aligned}
&\min_{y_{uv}} \sum_{(f_j,s_i) \in \overline{E}} c_{f_j,s_i} y_{f_j,s_i}\\
&\sum_{s_j} y_{0,s_j} = M,\\
&y_{s_i,f_i} - y_{0,s_i} - \sum_{f_j:\, (f_j,s_i) \in \overline{E}} y_{f_j,s_i} = -1,\ i \in \mathcal{J},\\
&\sum_{s_i:\, (f_j,s_i) \in \overline{E}} y_{f_j,s_i} + y_{f_j,2J+1} - y_{s_j,f_j} = 1,\ j \in \mathcal{J},\\
&\sum_{f_j} y_{f_j,2J+1} = M,\\
&0 \le y_{uv} \le 1,\ (u,v) \in E.
\end{aligned} \tag{3}$$

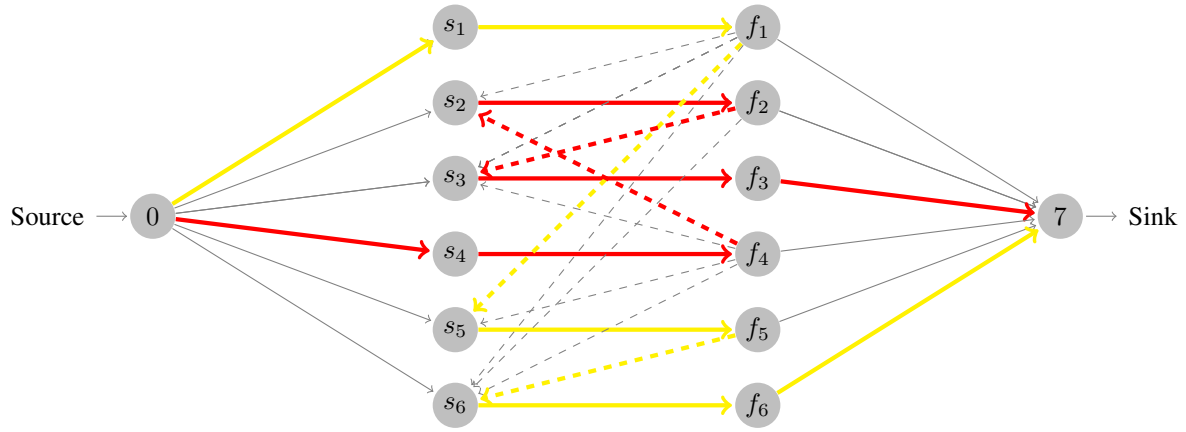


Figure 2 Operational FIS reformulation: network structure and the flow corresponding to the right schedule in Figure 1

The first constraint represents flow out of the source. The second constraints correspond to all starting nodes, whereas the third ones to the finishing nodes. The last constraint belongs to the sink. An important property of the constraints is that the corresponding matrix is totally unimodular, cf. [16], i.e. the integral choice of right-hand side vectors results into an integral solution. This enables to solve the problem by a linear programming technique to obtain an integral flow which identifies the assignment of the jobs to the machines. Figure 2 shows the network structure together with the feasible flow corresponding to the right schedule in Figure 1. As we discussed above, there are nodes for all starting and finishing times as well as artificial nodes for a source and sink. The unit flow goes through the nodes corresponding to the jobs which are assigned to the same machine. Different machines correspond to different flows marked by different colors (yellow and red). The costs are assigned to the employed dashed edges. Since job 2 follows after job 4 on the red machine, the dashed edge $\{f_4, s_2\}$ is active. Note that Ahuja et al. [2] provide a deep review of properties and algorithms for the network flow problems.

λ	min	$q_{0.25}$	$q_{0.50}$	$q_{0.75}$	max
0.2	13	15	15	16	18
0.5	18	20	21	21	24
0.8	19	22	23	24	28

Table 1 Descriptive statistics for the optimal numbers of machines in the tactical problem with $\varepsilon = 0.1$

4 Numerical study

In this section, we apply the proposed approach to simulated FIS instances with 100 jobs. In particular, we used independent exponential distributions to simulate the job lengths and the breaks between the jobs. We considered one choice of the parameter $\gamma = 0.2$ for the job length and three different choices for the breaks $\lambda \in \{0.2, 0.5, 0.8\}$, i.e. we are decreasing the expected length of the breaks. For each combination, we simulated 1000 problems. All simulations and computations were performed using Matlab R2016b. We prescribe the minimal acceptable reliability to 0.9, i.e. we set $\varepsilon = 0.1$. The delay distributions are characterized by the following probability distribution functions:

$$F_j(d) = P_j(D_j(\xi) \leq d) = 0.95 + 0.05(1 - e^{-0.2d}),\ d \geq 0, \tag{4}$$

i.e. there is probability 0.95 that the job j is finished in time and the conditional distribution of the delay length is exponential with the parameter equal to 0.2.

The minimal number of machines which is necessary to process all jobs is obtained using the left-edge algorithm, where the jobs are sorted according to their starting times and then sequentially assigned to the first available machine. The highest number of considered machines was set to 25 which was sufficient for all FIS instances to obtain the reliability at least 0.9.

Figures 3, 4, 5 show the highest attainable reliability for a particular problem instance and the boxplots based on 1000 simulated FIS problems. Table 1 contains the descriptive statistics for the minimal numbers of machines

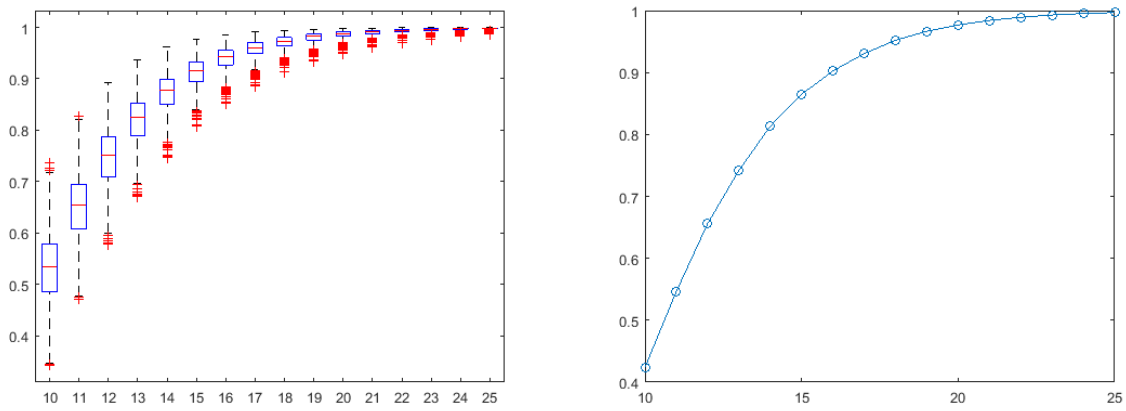


Figure 3 The highest attainable reliabilities for a given number of machines ($\lambda = 0.2$): 1000 instances (left), one instance (right)

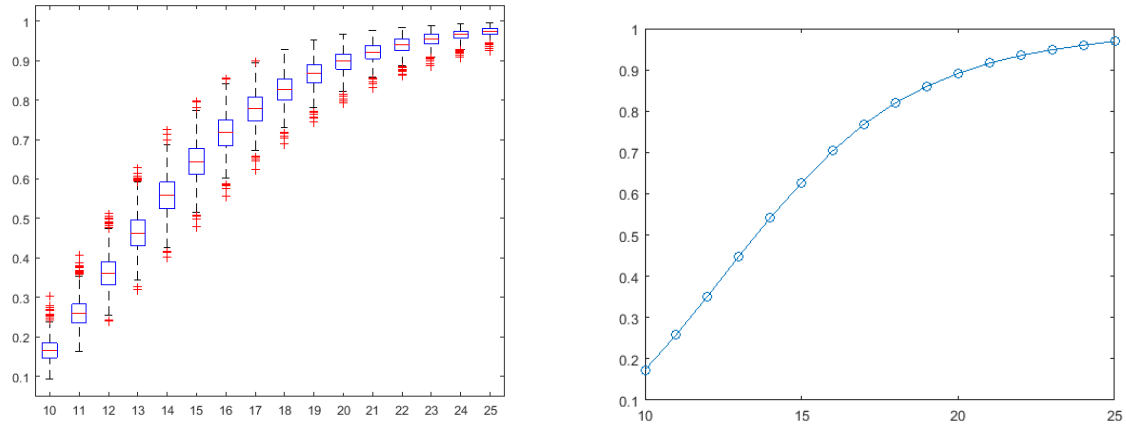


Figure 4 The highest attainable reliabilities for a given number of machines ($\lambda = 0.5$): 1000 instances (left), one instance (right)

which are necessary to reach the reliability at least 0.9, where q_α denotes the α quantile. We can observe that with decreasing the break lengths the necessary number of machines increases. The maximal number of required machines for $\lambda = 0.2$ is equal to the minimal number for $\lambda = 0.5$. The average computational time to solve one min-cost network flow problem was 2 seconds. Thus solving a tactical problem by evaluating 16 operational problems (for fixed number of machines between 10 and 25) takes less than 35 seconds in average.

5 Conclusions

We have investigated the tactical fixed interval scheduling problem under stochastic delays which we have formulated as a stochastic integer programming problem with a chance constraint. We have suggested a solution strategy based on solving a sequence of operational problems. The stochastic operational problem can be reformulated as a min-cost network flow problem and solved efficiently using, e.g., a linear programming solver. This approach has been employed in the numerical study showing a good performance on a large number of simulated instances. In the future research, we would like to focus on obtaining efficient schedules using the methods introduced by Branda and Kopa [5, 6], and Červinka et al. [8] .

Acknowledgements

This work was supported by the Czech Science Foundation under the Grant P402/12/G097 (DYME – Dynamic Models in Economics).

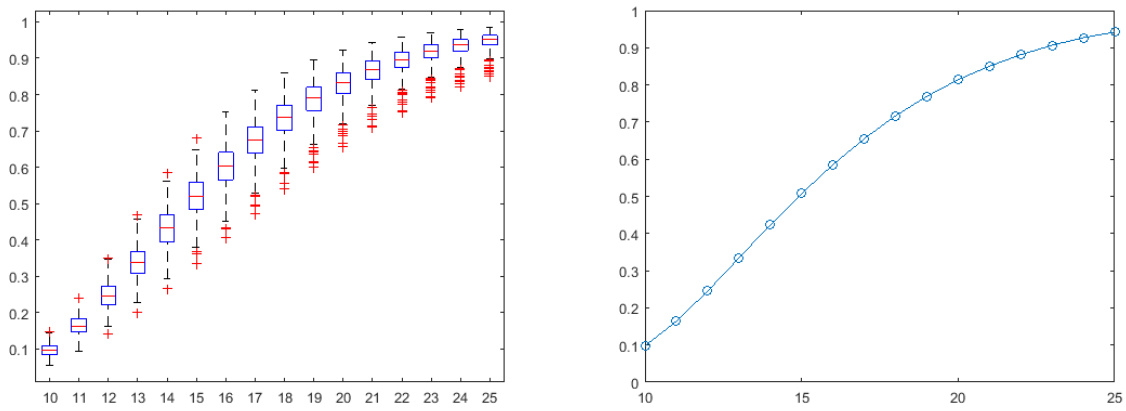


Figure 5 The highest attainable reliabilities for a given number of machines ($\lambda = 0.8$): 1000 instances (left), one instance (right)

References

[1] Adam, L., Branda, M.: Nonlinear chance constrained problems: optimality conditions, regularization and solvers, *Journal of Optimization Theory and Applications* **170** (2) (2016), 419–436.

[2] Ahuja, R.K., Magnanti, T.L., Orlin, J.B.: Network flows: theory, algorithms, and applications. Prentice-Hall, Inc., Upper Saddle River, NJ, 1993.

[3] Branda, M.: Distributionally robust fixed interval scheduling on identical machines. Submitted (2017).

[4] Branda, M., Hájek, Š.: Flow-based formulations for operational fixed interval scheduling problems with random delays. *Computational Management Science* **14** (1) (2017), 161–177.

[5] Branda, M., Kopa, M.: On relations between DEA-risk models and stochastic dominance efficiency tests. *Central European Journal of Operations Research* **22** (1) (2014), 13–35.

[6] Branda, M., Kopa, M.: DEA models equivalent to general N-th order stochastic dominance efficiency tests. *Operations Research Letters* **44** (2) (2016), 285–289.

[7] Branda, M., Novotný, J., Olstad, A.: Fixed interval scheduling under uncertainty – a tabu search algorithm for an extended robust coloring formulation. *Computers & Industrial Engineering* **93** (2016), 45–54.

[8] Červinka, M., Kanzow, C., Schwartz, A.: Constraint qualifications and optimality conditions for optimization problems with cardinality constraints. *Mathematical Programming* **130** (1) (2016), 353–377.

[9] Eliiyi, D.T.: Integrating tactical and operational decisions in fixed job scheduling. *Engineering Optimization* **45** (12) (2013), 1449–1467.

[10] Kolen, A.W., Lenstra, J.K., Papadimitriou, C.H., Spieksma, F.C.: Interval scheduling: a survey. *Naval Research Logistics* **54** (5) (2007), 530–543.

[11] Kovalyov, M.Y., Ng, C., Cheng, T.E.: Fixed interval scheduling: models, applications, computational complexity and algorithms. *European Journal of Operational Research* **178** (2) (2007) 331–342.

[12] Kroon, L.,Salomon, M.,Van Wassenhove, L.: Exact and approximation algorithms for the tactical fixed interval scheduling problem. *Operations Research* **45** (1997), 624–638.

[13] Ng, C.T., Cheng, T.C.E., Bandalouski, A.M., Kovalyov, M.Y., Lam, S.S.: A graph-theoretic approach to interval scheduling. *Journal of the Operational Research Society* **65** (10) (2014), 1571–1579.

[14] Smet, P., Wauters, T., Mihaylov, M., Berghe, G.V.: The shift minimisation personnel task scheduling problem: a new hybrid approach and computational insights. *Omega* **46** (2014), 64–73.

[15] Shapiro, A., Dentcheva, D., Ruszczyński, A.: *Lectures on stochastic programming: Modeling and theory*. SIAM, Philadelphia, 2009.

[16] Wolsey, L.A.: Integer programming. Wiley-Interscience series in discrete mathematics and optimization, J. Wiley & sons, New York (N.Y.), Chichester, Weinheim, 1998.

[17] Zhou, S., Zhang, X., Chen, B., van de Velde, S.: Tactical fixed job scheduling with spread-time constraints. *Computers & Operations Research* **47** (2014), 53–60.

Modelling synergy of the complexity and criticalness factors in the Project management

Helena Brožová, Jan Rydval[1]

Abstract. Project management is a discipline dealing with the management of a project team to achieve all project goals described by various specific project criteria as effectively as possible. Contemporary increasing of complexity of projects caused by many factors creates dangerous vulnerabilities. Some of these factors depend on the criticalness of the project activities also. Generally, the complexity and criticalness would be based on project weak spots and need to be evaluated by different factors. Very often, the synergy occurs between these factors.
In this paper, we suggest some approaches how to model the synergy effect. Synergy is the creation of a whole that is greater than the simple sum of its parts. Hence, the simple principle of weighted sum is not appropriate model. Contrary the multiplicative model does not give reasonable results in the case of zero evaluation of at least one factor. Therefore, four possible normalization of criticalness factors are suggested and simple additive weighted model and multiplicative model are compared.

Keywords: Project management, complexity, criticalness, threats, synergy modelling.

JEL Classification: O22, C44
AMS Classification: 91B50, 90B99

1 Introduction

Project management is a scientific and managerial discipline dealing with all phases of project live cycle as initiating, planning, executing, controlling, and closing with the aim of managing a project team to achieve all project goals described by various specific project criteria as effectively as possible. The identification of project threats and evaluation of the relevant project risks are its crucial part. The project threats and tisks arise from the complexity and criticalness of projects and create dangerous vulnerabilities. They can be caused by many factors, both in the environment or constraints of the project and in the choices made in achieving its goals (Goff, 2012, Williams, 1999). The complexity and criticalness of projects need good project risk management (Drahý, Pastor, 2016), which means the identification, measurement and prioritization of projects threats factors with efficient application of resources to minimize, monitor, and control the probability and impact of risk events.

Many authors endeavour to incorporate more factors into quantitative methods of project management. However they do not consider explicitly synergy effect of group of factors. Cooke-Davis (2002) draws on research from more than 70 organizations to identify factors that are critical to project success. Fatemi Ghomi and Teimouri (2002) and Madadi and Iranmanesh (2012) present overview on some indices (mainly for stochastic networks) from the literature and proposed a new one for quantification of the activity importance. Mota and Almeida (2011) proposes a multiple criteria decision model based on the ELECTRE TRI-C method to assign activities in project management to priorities classes for helping project managers to focus on the proper activities to ensure a successful project realization. Vidal et al. (2011) and Vidal and Marle (2015) define factors of project complexity and propose a multi-criteria approach to project complexity evaluation through the use of the Analytic Hierarchy Process (AHP).

The global evaluation of a set of factors of the complexity and criticalness are usually based on additive measure concepts which cannot, by definition, express the relationships of reinforcement or synergy which exist between considered factors

$$V(v_1,\ldots,v_p)=\sum_p w_p v_p \tag{1}$$

The application of additivity is based on the hypothesis of the interchangeability of the value of the different factors, seems intuitively justified. However, this method of calculation is irrelevant in the case of the structured and coherent set of assets which makes up a synergy. The synergy is based on the fact that the whole is more than the sum of its part. Therefore the synergic relation can be modelled by operators of non-additive integration (Bry, Casta, 2003) with following property

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems Engineering, Kamýcká 129, 165 21 Praha 6 – Suchdol, brozova@pef.czu.cz, rydval@pef.czu.cz

$$V(v_1,...,v_p) \geq \sum_p w_p v_p \tag{2}$$

This article builds on the previous articles Brožová et al. (2016, 2013) and develops the possibilities of modelling the synergy of the criticalness factors of the project activities, which results in overall criticalness potential. In the theoretical part, the authors suggest the suitable methods of normalization and modelling of synergy. In the second part, the example shows its application.

2 Modelling of the synergy of criticalness factors

The term synergy means "working together". The effect of an interaction of elements is greater than the simple sum of its action. Synergy effect is very important in the evaluation of criticalness potential of project activities. In managerial practice and especially in project management, synergy effect of key parameters of project successful completion must be taken into account. For instance, efficient resource allocation to individual project tasks and activities is not easy and, without its analysis, the critical resources can be firstly used for activities where synergy effect are not present. This can endanger the successful realization of the project.

The evaluation of project activities criticalness potential needs the suitable synthesis of partial evaluation using multi-attribute utility method. This approach seems to be more appropriate because it provides a solution in a clearer form describing the activities ranking. The managers accept better such type of results of complex problems. The application of multi-attribute utility method needs a good method of normalization or utility calculation and good way of synthesis of partial evaluation (Hwang, Yoon, 1981).

2.1 Normalization of criticalness identificators values

Criticalness factors are not expressed with the same standard, units (e.g., time, work-days, cost, probability etc. in the same time) and they cannot be simply composed into global criticalness potential. Normalization serves the purpose of bringing the all values into the same unit and the composing them into one value (Tofallis, 2014). Often used a general method of normalization is von Neumann and Morgenstern utility function converting the best value to 1and the worst observed value to zero. If this transformation is proportional, the risk neutral utility function is received. For the normalization of the criticalness identificators, the most critical values have to be transformed to the highest value to show the highest criticalness. The following four types of normalization are discussed and **Table 1** and **Figure 1** shows results of the following normalization forms.

1. The first type is range normalization (Tofalis, 2014) based on proportional linear 0-1 utility function, which represents risk neutral project manager normalization of criticalness factors.

$$u_I = \frac{\min\limits_K v_K - v_I}{\min\limits_K v_K - \max\limits_K v_K} \tag{3}$$

where v_I, v_K are criticalness evaluations (max value means the highest criticalness) and u_I is normalized value. The problem of this normalization method is, that at least one transformed value is always equal to 0 (the worst value). Therefore we suggest to use the following three normalization methods with values from $[1,10]$

2. The second normalization type is linear 1-10 normalization. It also represents risk neutral project manager position, but the lowest value is 1 and the highest is transformed to 10.

$$u_I = 9\frac{\min\limits_K v_K - v_I}{\min\limits_K v_K - \max\limits_K v_K} + 1 \tag{4}$$

where v_I, v_K are criticalness evaluations (max value means the highest criticalness) and u_I is normalized value.

3. The third type is exponential normalization represents value transformation suitable for risk prone project manager, because only highly critical evaluation is normalized into value near to 10.

$$u_I = 10^{\frac{\min\limits_K v_K - v_I}{\min\limits_K v_K - \max\limits_K v_K}} \tag{5}$$

where v_I, v_K are criticalness evaluations (max value means the highest criticalness) and u_I is normalized value.

4. The fourth type of normalization is power 1-10 normalization represents risk averse project manager approach. Already in the case of a weak criticality, this evaluation is normalized to a high value (close to 10).

$$u_I = -9\left(\frac{\min\limits_K v_K - v_I}{\min\limits_K v_K - \max\limits_K v_K}\right)^2 + 18\left(\frac{\min\limits_K v_K - v_I}{\min\limits_K v_K - \max\limits_K v_K}\right) + 1 \tag{6}$$

where v_I, v_K are criticalness evaluations (max value means the highest criticalness) and u_I is normalized value.

	Initial values	0-1 normalization	1-10 normalization	EXP normalization	Power normalization
A	**0**	0	1	1	1
B	**0,1**	0,1	1,9	1,259	2,71
C	**0,2**	0,2	2,8	1,585	4,24
D	**0,3**	0,3	3,7	1,995	5,59
E	**0,4**	0,4	4,6	2,512	6,76
F	**0,5**	0,5	5,5	3,162	7,75
G	**0,6**	0,6	6,4	3,981	8,56
H	**0,7**	0,7	7,3	5,012	9,19
I	**0,8**	0,8	8,2	6,310	9,64
J	**0,9**	0,9	9,1	7,943	9,91
K	**1**	1	10	10	10

Table 1 Four types of normalization of criteria values

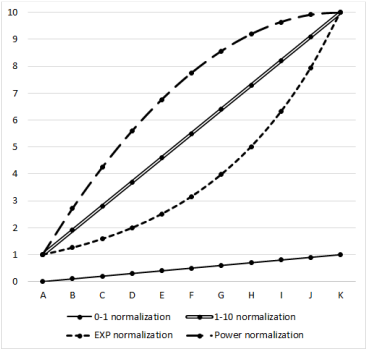


Figure 1 Graph of normalized values

2.2 Synthesis of criterial values

Synthesis of normalized partial criticalness factors values gives a global evaluation of project activity criticalness. It is again important to select the best form of this aggregation (Tofallis, 2014).Because the synergy effect is important in the interaction of partial criticalness factors, the comparison of synthesis based on simple additive weighting method and one based on the multiplicative method will be discussed.

Using the simple additive weighting method all criticalness factors evaluations are synthesized using formula

$$c_I = \sum_p w_p u_{Ip} \tag{7}$$

where c_I is criticalness potential of the activities, u_{Ip} is normalized evaluation of each components of activities criticalness, p is number of used components of criticalness and w_p is the weight of the p component of criticalness.

In the multiplicative model all criticalness evaluations are multiplied using formula

$$c_I = \prod_p u_{Ip} \tag{8}$$

where c_I is criticalness potential of the activities, u_{Ip} is evaluation of each components of the criticalness and p is number of used components of criticalness.

Table 2 and **Figure 2** shows the hypothetical results of synthesis of three criticalness values using the multiplicative model. It is possible to see the synergy effect. The multiplicative aggregation shows nonlinear dependence between the partial factors values and their synthesised value. More is even more and less is even less.

	Initial values			0-1 normalization	1-10 normalization	EXP normalization	Power normalization
A	0	0	0	0	1	1	1
B	0,1	0,1	0,1	0,001	6,859	1,995	19,903
C	0,2	0,2	0,2	0,008	21,952	3,981	76,225
D	0,3	0,3	0,3	0,027	50,653	7,943	174,677
E	0,4	0,4	0,4	0,064	97,336	15,849	308,916
F	0,5	0,5	0,5	0,125	166,375	31,623	465,484
G	0,6	0,6	0,6	0,216	262,144	63,096	627,222
H	0,7	0,7	0,7	0,343	389,017	125,893	776,152
I	0,8	0,8	0,8	0,512	551,368	251,189	895,841
J	0,9	0,9	0,9	0,729	753,571	501,187	973,242
K	1	1	1	1	1000	1000	1000

Table 2 Multiple aggregation of three criteria values

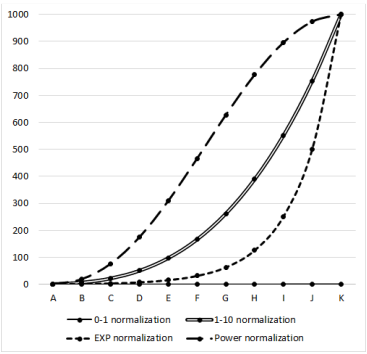


Figure 2 Graph of synthesized values

3 Practical application of synergy modelling

The synergy modelling and analysis of derived indicators of project activity criticalness is described on the following small-scale project (Brožová et al., 2013) and compare with their simple addition. A critical path of a project (**Figure 3**) is composed of activities A, C, D, F, I, J, K and N. Individual criticalness factors give an information on how project activities potentially endanger a successful realisation of a project from a different point of view. Estimation of overall criticalness of the project activities is based on the multiple criteria decision making methods using five indicators of the criticalness. We use five criticalness indicators for all activities; specifically, the duration of the activity (time criticalness), its topological location in the structure of the project activities (topological criticalness), the time reserve of the activity (slack criticalness), the activity costs (cost

criticalness) and the activity work (work criticalness). The real values of criticalness indicators and their 0-1normalization are in the **Table 3**.

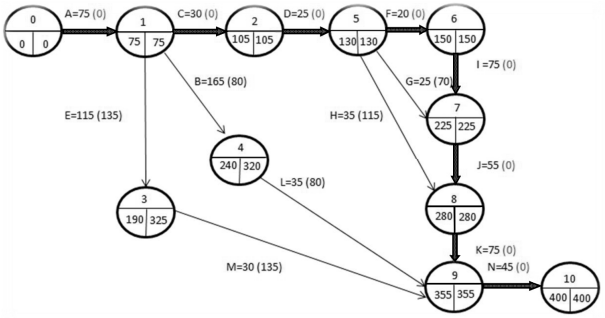


Figure 3 Small-scale project with the indication of a critical path

Activity	Days	Time criticalness	Probability of critical path	Topological criticalness	Slack	Slack criticalness	Cost	Cost criticalness	Work	Work criticalness
A	75	0.379	1	1	0	1	1350	0.656	450	0.198
B	165	1	0.33	0.247	80	0.407	990	0.473	495	0.221
C	30	0.069	0.33	0.247	0	1	630	0.290	630	0.290
D	25	0.034	0.33	0.247	0	1	175	0.059	175	0.059
E	115	0.655	0.33	0.247	135	0	690	0.321	345	0.145
F	20	0	0.11	0	0	1	60	0	60	0
G	25	0.034	0.11	0	70	0.481	150	0.046	150	0.046
H	35	0.103	0.11	0	115	0.148	210	0.076	210	0.076
I	75	0.379	0.11	0	0	1	1575	0.771	1575	0.771
J	55	0.241	0.22	0.124	0	1	1155	0.557	1155	0.557
K	75	0.379	0.33	0.247	0	1	2025	1	2025	1
L	35	0.103	0.33	0.247	80	0.407	875	0.415	875	0.415
M	30	0.069	0.33	0.247	135	0	360	0.153	180	0.061
N	45	0.172	1	1	0	1	945	0.450	945	0.450

Table 3 Activities evaluation, 0-1 criticalness indicators, input values for criticalness potential evaluation

The additive model with weights of used indicators (Brožová et al., 2013) then shows the overall criticalness potential of each activity and its actual threat of the successful realisation of the project. Some activities, which do not lie on the critical path, can be regarded as a greater threat to the project than activities which lie on the critical path. Several activities have much higher criticalness potential towards the project than the critical path method reveals. Then it is necessary to pay close attention to those activities with the highest criticalness and thus preventing project failure (**Table 4**, **Figure 4**).

0-1 normalization	Time criticalness	Topological criticalness	Slack criticalness	Cost criticalness	Work criticalness	Multiplicative model	Rank	Additive model with weights	Rank
K	0,379	0,247	1	1	1	0,094	1	0,756	1
A	0,379	1	1	0,656	0,198	0,049	2	0,615	2
N	0,172	1,000	1	0,450	0,450	0,035	3	0,580	4
I	0,379	0	1	0,771	0,771	0	9	0,590	3
J	0,241	0,124	1	0,557	0,557	0,009	5	0,481	5
B	1	0,247	0,407	0,473	0,221	0,011	4	0,451	6
C	0,069	0,247	1	0,290	0,290	0,001	7	0,337	7
L	0,103	0,247	0,407	0,415	0,415	0,002	6	0,331	8
E	0,655	0,247	0	0,321	0,145	0	9	0,280	9
D	0,034	0,247	1	0,059	0,059	0,0001	8	0,212	10
M	0,069	0,247	0	0,153	0,061	0	9	0,116	12
G	0,034	0	0,481	0,046	0,046	0	9	0,092	13
H	0,103	0	0,148	0,076	0,076	0	9	0,076	14
F	0	0	1	0	0	0	9	0,129	11
Weights	0,164	0,189	0,129	0,288	0,230				

Table 4 0-1 normalization of criticalness values of activities and their synthesis

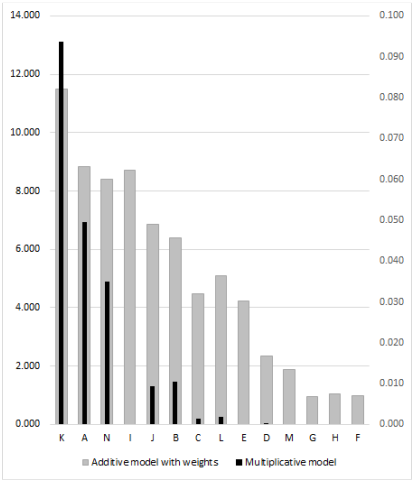


Figure 4 Synthesized 0-1 values

This effect is much more important if the synergy of criticalness factors is supposed. In this case, the multiplicative approach is more suitable. Unfortunately, if some normalized values of criticalness indicators are equal to 0, this approach fails. For instance, the activity I, which is critical according to the CPM method has the value of multiplicative criticalness potential equal to 0. Therefore, three other possible normalizations of criticalness factors are used and weighted sum and multiplicative model are compared.

The normalization type 1-10 (**Table 5**, **Figure 5**) give reasonable results for risk neutral managers. Multiplicative approach represents an appropriate model of synergy effect as **Figure 5** shows. This type of normalization provides a clear synergic effect of criticality factors, especially for activity K. Activity K has the highest criticalness in 3 out of 5 factors, thus its criticalness potential is the highest of all activities and this activity threatens the success of the entire project. This result can be also achieved by the simple additive approach with weights, however, by the additive approach, the difference between the first and second activity with the highest criticalness potential is comparable to the differences between other activities. Multiplicative approach also shows differences among the second, third and fours activities, which have similar values of criticalness potential according to the additive approach.

1-10 normalization	Time criticalness	Topological criticalness	Slack criticalness	Cost criticalness	Work criticalness	Multiplicative model	Rank	Additive model with weights	Rank
K	4,414	3,225	10	10	10	14233,24	1	7,802	1
A	4,414	10	10	6,908	2,786	8495,93	2	6,536	2
N	2,552	10	10	5,053	5,053	6516,39	3	6,217	4
I	4,414	1	10	7,939	7,939	2781,87	4	6,314	3
J	3,172	2,112	10	6,015	6,015	2424,75	5	5,325	5
B	10	3,225	4,667	5,260	2,992	2368,43	6	5,055	6
C	1,621	3,225	10	3,611	3,611	681,35	7	4,036	7
L	1,931	3,225	4,667	4,733	4,733	650,92	8	3,979	8
E	6,897	3,225	1	3,885	2,305	199,21	9	3,519	9
D	1,310	3,225	10	1,527	1,527	98,49	10	2,906	10
M	1,621	3,225	1	2,374	1,550	19,23	11	2,045	12
G	1,310	1	5,333	1,412	1,412	13,94	12	1,824	13
H	1,931	1	2,333	1,687	1,687	12,82	13	1,681	14
F	1	1	10	1	1	10	14	2,162	11

Table 5 1-10 normalization of criticalness values of activities and their synthesis

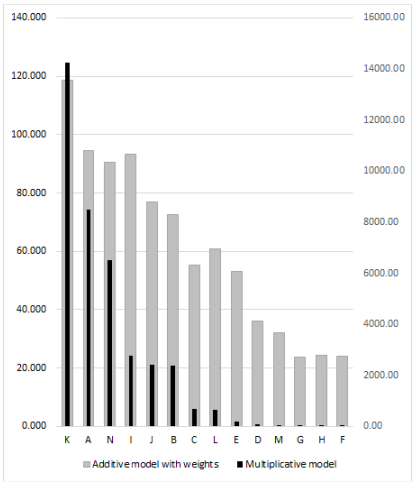


Figure 5 Synthesized 1-10 values

The exponential type of normalization (**Table 6**, **Figure 6**) gives reasonable results for risk prone managers (only higher value of criticalness indicators or potential is seen as really threatening). A small value of criticalness factor is transformed into a smaller value of criticalness indicator. Higher values of criticalness indicators are only reached if criticalness factors have really high values. Activity K again has the highest criticalness potential because its criticality in 3 out of 5 factors is the highest. This activity now is incomparable with others. In the same moment, the difference among other activities occurs.

EXP normalization	Time criticalness	Topological criticalness	Slack criticalness	Cost criticalness	Work criticalness	Multiplicative model	Rank	Additive model with weights	Rank
K	2,395	1,767	10	10	10	4231,568	1	7,195	1
A	2,395	10	10	4,534	1,579	1715,032	2	5,244	2
N	1,487	10	10	2,821	2,821	1183,524	3	4,887	4
I	2,395	1	10	5,902	5,902	834,248	4	4,928	3
J	1,743	1,329	10	3,608	3,608	301,633	5	3,696	5
B	10	1,767	2,555	2,974	1,665	223,488	6	3,544	6
C	1,172	1,767	10	1,950	1,950	78,761	7	2,827	7
L	1,269	1,767	2,555	2,599	2,599	38,687	8	2,217	9
E	4,520	1,767	1	2,092	1,396	23,335	10	2,128	11
D	1,083	1,767	10	1,144	1,144	25,045	9	2,395	8
M	1,172	1,767	1	1,421	1,151	3,388	13	1,329	13
G	1,083	1	3,030	1,111	1,111	4,051	12	1,333	12
H	1,269	1	1,407	1,192	1,192	2,537	14	1,196	14
F	1	1	10	1	1	10	11	2,162	10

Table 6 EXP normalization of criticalness values of activities and their synthesis

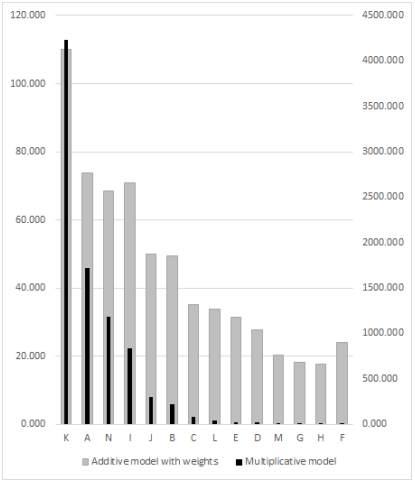


Figure 6 Synthesized EXP values

Power normalization (**Table 7**, **Figure 7**) compared to the previous two types of normalization (1-10 and EXP normalization) represents the risk averse managers (only small value of criticalness indicators or potential is seen as really non-threatening). For instance the activity I is evaluated as less threatening than the activity B. The disadvantage of this normalization is the magnitude of the result, which is often difficult to analyse or comprehend mainly in the case of many partial criticalness indicators aggregation.

	Powe normalization	Time criticalness	Topological criticalness	Slack criticalness	Cost criticalness	Work criticalness	Multiplicative model	Rank	Additive model with weights	Rank
K		6.533	4.9	10	10	10	32007.0	1	127.764	1
A		6.533	10	10	8.938	4.218	24628. 6	2	114.627	2
N		3.836	10	10	7.281	7.281	20336.8	3	113.203	3
I		6.533	1	10	9.528	9.528	5930. 6	6	111.817	4
J		4.820	3.09	10	8.236	8.236	10094.1	5	103.025	5
B		10	4.9	6.84	7.503	4.544	11424.2	4	97.228	6
C		2.199	4.9	10	5.464	5.464	3216.1	8	78.272	8
L		2.766	4.9	6.84	6.917	6.917	4434.9	7	89.326	7
E		8.930	4.9	1	5.846	3.421	875.1	9	76.473	9
D		1.610	4.9	10	2.023	2.023	322.7	10	46.122	10
M		2.199	4.9	1	3.538	2.066	78.7	11	46.050	11
G		1.610	1	7.58	1.806	1.806	39.8	13	30.050	13
H		2.766	1	3.469	2.322	2.322	51.7	12	32.895	12
F		1	1	10	1	1	10	14	24.000	14

Table 7 Power normalization of criticalness values of activities and their synthesis

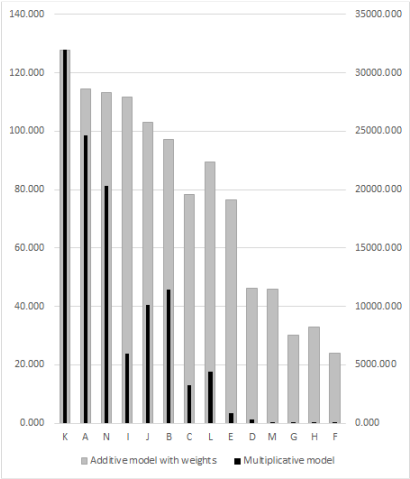


Figure 7 Synthesized power values

Conclusion

This article compares four possibilities of modelling the synergy effect of the criticalness factors of the project activities, which results in overall criticalness potential. The synergy effect has nonlinear character, its growth rate increases with the higher evaluation and with the number of contributing factors. Therefore the multiplicative approach to synthesis of the partial criticalness factors highlights the differences between the criticalness potential of activities. This showed the project activities that need to be targeted in order to achieve the successful realization of the project. For activities with a high criticalness potential, careful time monitoring, the allocation of critical resources or costs is necessary. The appropriate type of normalization with a multiplicative approach to calculate the criticalness potential values is a good way to identify highly critical project activities. By using this approach, it is possible to achieve a greater success of project realizations.

References

[1] Brožová, H., Bartoška, J., Šubrt, T., Rydval, J.: Multiple Criteria Evaluation of the Activities Criticalness in the Project Management. In: *Proceedings of the 31th Conference on Mathematical Methods in Economics.* VŠP Jihlava, 2013. 83-89.

[2] Brožová, H., Bartoška, J., Šubrt, T., Rydval, J.: Task Criticalness Potential: A Multiple Criteria Approach to Project Management. *Kybernetika*, **52 (**2016), 558-574.

[3] Bry, X., Casta, J.F.: Synergy modelling and financial valuation: the contribution of Fuzzy Integrals. *Connectionist Approaches in Economics and Management Sciences* **6** (2003) pp 165-182.

[4] Cooke-Davies, T.: The "real" success factors on projects. *International Journal of Project Management* **20**, 2002, 185–190.

[5] Drahý, C., Pastor, O.: Relationship between Project Complexity and Risk Kinds Identified on Project, *Economics, Management, Inovations*, **8** (2016), 42-50.

[6] Fatemi Ghomi, S. M. T., Teimouri, E.: Path critical index and activity critical index in PERT networks. *European Journal of Operational Research* **141** (2002), 147-152.

[7] Goff, S.A.: Revisiting Risks: Threats and Opportunities in Complex Projects. *Proceedings of the 6th Annual Project Management Symposium*, University of Texas, Dallas, 2012.

[8] Hwang, Ch-L., Yoon, K.: *Multiple Attribute Decision Making*. Springer Verlang Berlin Heidelberg, New York, 1981.

[9] Madadi, M., Iranmanesh, H.: A management oriented approach to reduce a project duration and its risk (variability). *European Journal of Operational Research* **219** (2012), 751-761.

[10] Mota, C. M. M., de Almeida, A. T.: A multi-criteria decision model for assigning priority classes to activities in project management. *Ann Oper Res* **199** (2011), 361-372.

[11] Tofallis, C.: Add or Multiply? A Tutorial on Ranking and Choosing with Multiple Criteria. *INFORMS Transactions on Education* **14** (2014), 109-119.

[12] Vidal, L.A., Marle, F.: Understanding project complexity: implications on project management. *Kybernetes*, Emerald, 2015, <10.1108/03684920810884928>. <hal-01215364>.

[13] Vidal, L.A., Marle, F., Bocquet, J.C.: Measuring project complexity using the Analytic Hierarchy Process, *International Journal of Project Management* **29** (2011), 718–727.

[14] Williams, T. M.: The need for new paradigms for complex projects. *International Journal of Project Management* **17** (1999), 269-273.

Inflation targeting and variability of money market interest rates under a zero lower bound

Karel Brůna[1], Quang Van Tran[2]

Abstract. The paper presents a formal framework of money market interest rates variability under a zero lower bound in a monetary policy strategy of inflation targeting. The potential factors influencing the variability of money market interest rates are considered within a near zero level of main policy rate variability. At the same time, model optimal main policy rate shows significant volatility due to changes in the structural characteristics of economy facing deep economic and financial shocks, changing perception of inflation risks, central bank's weakened credibility and uncertainty about the efficient transmission of monetary measures. Money market interest rates are modeled with a VAR model with exogenous variables. The model parameters are estimated with the use of publicly available quarterly data for the Czech Republic for a period from 2000q1 to 2016q4 in two stages. First the parameters of the corresponding model are identified using the Bayesian technique. Then the matrix used to convert the structural system into its reduced form is quantified through the sign restriction approach. These two parameter groups are then used to decompose money market interest rate series into a series of cumulative structural shocks of each endogenous component in the model.

Keywords: Money market interest rates, volatility, inflation targeting, Bayesian VAR, Historical decomposition

JEL Classification: E43-4
AMS Classification: 91G70

1 Introduction

Global financial and economic crisis caused a decline in monetary policy interest rates and money market interest rates close to zero level and in some cases even slightly below zero. Alternative monetary policy measures were being applied to overcome restrictive character of expected deflation, high credit risk and debt deleveraging resulting in high expected real interest rates, high risk premiums and high debt burden. The paper presents formal framework of money market interest rates variability under a zero lower bound in a monetary policy strategy of inflation targeting. The potential factors influencing the variability of money market interest rates are considered within a near zero level of main policy rate volatility when model optimal main monetary policy rate shows significant volatility due to changes in the structural characteristics of economies. To quantify the variability of money market rate around zero bound, we use a Vector Auto Regressive (VAR) model with exogenous variables estimated by Bayesian technique for its reduced form. The corresponding structural form is then recovered by sign restriction method. The obtained coefficients are used to decompose residuals into the corresponding structural shocks which allows us to express money market interest rate as a function of individual endogenous structural shocks, hence to measure how each of them contributes to total variability of money market interest rate. For estimation purpose, we use quarterly data from period 2000q1 to 2016q4 from public available sources.

2 Money market yield curve variability and inflation targeting

A concept of preferred habitat theory assumes that agents in the money market compare actual *n*-day interest rate with the expected future development of O/N interest rates average level in *n*-day horizon. The preferences of the agents also make them request increasing term premium reflecting higher maturity included risk premium as a price for credit risk. Therefore, the money market equilibrium can be in general expressed as a position of a risk averse speculator who (based on available information (Ω_t)) quotes the actual *n*-day interest rate (IR_t^n) as the sum of the expected average of O/N rates in the period of *t* to *t+k* and term premium (ϕ_t^n):

[1] University of Economics, Prague, Department of Monetary Theory and Policy, W. Churchill sq. 4, Prague, Czech Republic e-mail: bruna@vse.cz.

[2] University of Economics, Prague, Department of Monetary Theory and Policy, W. Churchill sq. 4, Prague, Czech Republic e-mail: tran@vse.cz.

$$IR_t^n = \frac{1}{n}\sum_{k=0}^{n-1} IR_{t+k}^{O/N,e}\left|\Omega_t + \phi_t^n\right. \quad (1)$$

In a framework of inflation targeting, it is assumed that central bank uses positive *s*-day targeted interest rate ($IR_{CB,t}^{TARGET,s}$) as a price of main policy operation while the maturity of this rate maximally equals the maturity of ultra-short money market rates. Central bank's main policy operations are typically used to manage the liquidity of the banking system within a minimum reserves maintenance period. It is assumed that the duration of maintenance period *r* is longer/maximally equals the maturity of main policy rate (i.e. $s \leq r$) while it holds good that *r* is usually the whole multiple of *s.* The end of the maintenance period would then be identical with the maturity of a tender for liquidity supply/withdrawal at *s*-day main policy rate.

The decision of central bank on setting the main policy rate is a problem of optimal management of an economic system. The main interest rate is considered as an instrumental variable that minimises the values of future deviations of the variables that are the object of central bank´s interest from targeted values in relation to restrictions given by the existing structure of economy. Thus the loss function (L) of central bank is usually the sum of the present values of quadratic deviations of expected inflation (π_{t+i}^e) and expected economic growth (y_{t+i}^e) from inflation target (π_{t+i}^{TARGET}) and potential output (y_{t+i}^*) (Srour, 1999):

$$L_t = \sum_{i=0}^{\infty} \delta^i \left[(1-\alpha)(\pi_{t+i}^e - \pi_{t+i}^{TARGET})^2 + \alpha(y_{t+i}^e - y_{t+i}^*)^2\right] \quad (2)$$

where the discount factor δ is usually considered as $0 \leq \delta \leq 1$, α and (1-α) are the relative weights of quadratic deviations of inflation and economic growth from targeted values and *i* is a time shift indicating forward looking character of monetary policy and the existence of time lags of the effects within transmission mechanism.

To set the main policy rate the present economic structure implies restrictions for achievement of monetary targets. With regard to the central bank's loss function and model of economic structure, the main interest rate that minimises central bank's loss function for the actual inflation forecast applied for the model of economic structure in the horizon of efficient transmission (*k* period) is considered as optimal. Formally the optimal level of the main interest rate ($IR_{CB,t}^{OPTIM}$) is defined by the reaction function of central bank of the Taylor´s type:

$$IR_{CB,t}^{OPTIM} = IR_{CB}^{EQ} + \beta_t(\pi_{t+k}^e - \pi_{t+k}^{TARGET}) + \gamma_t(y_{t+k}^e - y_{t+k}^*) \quad (3)$$

where IR_{CB}^{EQ} is the main policy rate corresponding to long-term economic equilibrium, parameters β_t and/or γ_t express the intensity by which the central bank reacts to the overshooting (undershooting) of inflation target and/or positive (negative) output gap. As stated by Svensson (2000) and Favero and Rovelli (2000), time variable parameters β_t and γ_t are the convolution of parameters describing central bank's preferences to the inflation and business cycle and structural characteristics of economy.

Sack (1998) point out that there exists a disproportion between the model optimum level of the main policy rate and the targeted main policy rate ($IR_{CB,t}^{TARGET}$) which is less volatile and strongly positively correlated with its lagged values (Clarida, Gali and Gertler 1998) to avoid impairment of central bank's credibility that could be connected with high variability of the main policy rate (Goodhart 1998). Therefore, the variability of the targeted main policy rate often becomes a part of the central bank's objective function while optimal main policy rate is based to achieve monetary-policy objectives under the lowest possible variability of the main interest rate. The dynamics of targeted main interest rate may show deviations from monetary rule (3) due to application of alternative monetary policy strategies that are presented as a correlation of random shocks w_t (Rudebusch 2002):

$$IR_{CB,t}^{TARGET} = (1-\rho_1-\rho_2)IR_{CB,t}^{OPTIM} + \rho_1 IR_{CB,t-1}^{TARGET} + \rho_2 w_t \quad (4)$$

$$w_t = \sum_{n=0}^{q} D_1 U_{1,t-n} + \sum_{n=0}^{r} D_2 U_{2,t-n} + \ldots + \sum_{n=0}^{z} D_m U_{m,t-n} \quad (5)$$

where ρ_1 is the weight of the lagged targeted main policy rate, being the measure of central bank's aggressiveness in the stabilisation of inflation and output gap, $D_1, D_2, \ldots, D_m$ are the vectors of parameters, $U_{1,t-n}, U_{2,t-n}, \ldots, U_{m,t-n}$ are the actual and lagged values of variables considered by the central bank for setting the main interest rate.

It is normally assumed that optimal main interest rate is positive high volatile interest rate that stabilizes inflation and output gap to (near) zero levels considering a model framework of an economy. In case of expected deep

deflation, economic recession and financial crisis it seems to be optimal to set the main policy rate to high negative level and to stay below zero for a long period to enable the real and financial sector to recover. When optimal main policy rate goes to negative, inflation targeting policy faces up its limitation as targeted main policy rate is rarely used to cross its zero lower bound (Bernanke 2000). A set of alternative monetary policies (outright purchases of assets, home currency depreciation, money-financed aggregate demand expansion) are typically suggested to overcome a restrictive character of near zero level targeted main policy rates. Using these alternative monetary policy measures, targeted main policy rate is meantime fixed to (near) zero level for a long time which reduces its variability and respective covariances to zero in the same period as well. This fact deals with an implicit change in monetary policy mechanism of inflation targeting within which unconventional monetary policy instruments take over a position of main interest rate in the reaction function. Therefore, variability of targeted main interest rate is considered as positive if and only if central bank believes that reaction function could be effectively applied to reach monetary policy targets in the central bank´s loss function. For the case of $IR_{CB,t}^{OPTIM} < 0$:

$$\begin{aligned} \operatorname{var} IR_{CB,t}^{TARGET} &= (1-\rho_1-\rho_2)^2 \operatorname{var} IR_{CB,t}^{OPTIM} + \rho_1{}^2 \operatorname{var} IR_{CB,t-1}^{TARGET} + \rho_2{}^2 \operatorname{var} w_t + \\ &+ 2(1-\rho_1-\rho_2)\rho_1 \operatorname{cov}(IR_{CB,t}^{OPTIM}, IR_{CB,t-1}^{TARGET}) + 2(1-\rho_1-\rho_2)\rho_2 \operatorname{cov}(IR_{CB,t}^{OPTIM}, w_t) + \\ &+ 2\rho_1\rho_2 \operatorname{cov}(IR_{CB,t-1}^{TARGET}, w_t) = 0 \end{aligned} \tag{6}$$

Zero variability of targeted main policy rate reflects the case when random shocks offset optimal main policy rate (high negative covariance between these variables is maintained) and weight ρ_2 is optimized according it:

$$\operatorname{var} w_t = \frac{-\left[(1-\rho_1-\rho_2)^2 \operatorname{var} IR_{CB,t}^{OPTIM} + 2(1-\rho_1-\rho_2)\rho_2 \operatorname{cov}(IR_{CB,t}^{OPTIM}, w_t)\right]}{\rho_2{}^2} \tag{7}$$

Although targeted main policy rate remains at a zero level which is typically maintained by overall excess of banking system liquidity provided by central bank through alternative monetary policy instruments, the dynamics of economic fundamentals behind money market interest rates could be a source of their significant nominal and real variability. Although the variability of targeted main policy rate is at a zero level, the variability of economic fundamentals is still reflected in the variability of optimal main policy rate, which using Goodman's breakdown of the product of random variables (Goodman 1960) could be written in the following form:

$$\begin{aligned} \operatorname{var} IR_{CB,t}^{OPTIM} &= \bar{\beta}^2 \operatorname{var}(\pi_{t+k}^e - \pi_{t+k}^{TARGET}) + \overline{(\pi_{t+k}^e - \pi_{t+k}^{TARGET})}^2 \operatorname{var}\beta + 2\bar{\beta}\overline{(\pi_{t+k}^e - \pi_{t+k}^{TARGET})} \operatorname{cov}_{11}^{\beta\pi} + \\ &+ 2\bar{\beta} \operatorname{cov}_{12}^{\beta\pi} + 2\overline{(\pi_{t+k}^e - \pi_{t+k}^{TARGET})} \operatorname{cov}_{21}^{\beta\pi} + \operatorname{var}\left[\Delta\beta\Delta(\pi_{t+k}^e - \pi_{t+k}^{TARGET})\right] + \\ &+ \bar{\gamma}^2 \operatorname{var}(y_{t+k}^e - y_{t+k}^*) + \overline{(y_{t+k}^e - y_{t+k}^*)}^2 \operatorname{var}\gamma + 2\bar{\gamma}\overline{(y_{t+k}^e - y_{t+k}^*)} \operatorname{cov}_{11}^{\gamma\gamma} + 2\bar{\gamma} \operatorname{cov}_{12}^{\gamma\gamma} + \\ &+ 2\overline{(y_{t+k}^e - y_{t+k}^*)} \operatorname{cov}_{21}^{\gamma\gamma} + \operatorname{var}\left[\Delta\gamma\Delta(y_{t+k}^e - y_{t+k}^*)\right] + 2\operatorname{cov}\left\{\left[\beta(\pi_{t+k}^e - \pi_{t+k}^{TARGET})\right], \left[\gamma(y_{t+k}^e - y_{t+k}^*)\right]\right\} \end{aligned} \tag{8}$$

where $\bar{\beta}$ and/or $\bar{\gamma}$ is the mean value of parameter β and/or γ, the term $\overline{(\pi_{t+k}^e - \pi_{t+k}^{TARGET})}$ and/or $\overline{(y_{t+k}^e - y_{t+k}^*)}$ is the mean value of the deviation of expected inflation from the inflation target and/or expected output gap, $\operatorname{cov}_{ij}^{\beta\pi}$ is $\operatorname{cov}\left[(\Delta\beta)^i, (\Delta(\pi_{t+k}^e - \pi_{t+k}^{TARGET}))^j\right]$, $\operatorname{cov}_{ij}^{\gamma\gamma}$ is $\operatorname{cov}\left[(\Delta\gamma)^i, (\Delta(y_{t+k}^e - y_{t+k}^*))^j\right]$, the symbol Δ expresses the deviation of the respective variable from its mean value (e.g. $\Delta\beta = \beta - \bar{\beta}$).

Under a zero lower bound, the variability of *n*-month money market interest rate is in general limited by negative covariance between expected optimal main policy rate and the other non-specified factors in central bank's reaction function. The variability of money market interest rates reflects the variability of risk and term premium that could be the main source of money market interest rate instability while monetary impulses are temporally (until the exit strategy is being applied) weak:

$$\begin{aligned} \operatorname{var} IR_t^n &= \frac{1}{n^2}\left[(1-\rho_1-\rho_2)^2 \operatorname{var}\sum_{j=0}^{n=1} IR_{CB,t+j}^{OPTIM,e} + \rho_2 \operatorname{var}\sum_{j=0}^{n=1} w_{t+j}^e + 2(1-\rho_1-\rho_2)\rho_2 \operatorname{cov}\left(\sum_{j=0}^{n=1} IR_{CB,t+j}^{OPTIM,e}, \sum_{j=0}^{n=1} w_{t+j}^e\right)\right] + \\ &+ \rho_2^2 \operatorname{var}\phi_t^n + \frac{2}{n}\operatorname{cov}\left\{\left[(1-\rho_1-\rho_2)\sum_{j=0}^{n=1} IR_{CB,t+j}^{OPTIM,e} + \rho_1\sum_{j=0}^{n=1} IR_{CB,t+j-1}^{TARGET,e} + \rho_2\sum_{j=0}^{n=1} w_{t+j}^e\right], \phi_t^n\right\} \end{aligned} \tag{9}$$

$$\operatorname{var}\sum_{j=0}^{n-1} IR_{CB,t+j}^{OPTIM,e} = \operatorname{var}\sum_{j=0}^{n-1}\beta_{t+j}(\pi_{t+k+j}^{e,t+j} - \pi_{t+k+j}^{TARGET}) + \operatorname{var}\sum_{j=0}^{n-1}\gamma_{t+j}(y_{t+k+j}^{e,t+j} - y_{t+k+j}^{*}) + \\ + 2\operatorname{cov}(\sum_{j=0}^{n-1}\beta_{t+j}(\pi_{t+k+j}^{e,t+j} - \pi_{t+k+j}^{TARGET}), \sum_{j=0}^{n-1}\gamma_{t+j}(y_{t+k+j}^{e,t+j} - y_{t+k+j}^{*})) \tag{10}$$

3 The effect of endogenous variables on money market interest rate

As stated before, the movement of money market interest rate is caused by changes (shocks) in endogenous and exogenous variables included in the model. To measure their impact, we use the historical decomposition technique for a VAR model with exogenous variables. The gist of this technique is briefly described below.

3.1 Historical decomposition

For our econometric analysis we use a VAR model with exogenous variables in the following structural form:

$$DY_t = C + \sum_{i=1}^{p} A_i Y_{t-i} + BZ_t + U_t\,, \tag{11}$$

where Y_t is a vector of endogenous variables, Z_t is a vector of exogenous variable, U_t is a vector of structural shocks, *C* is vector of constants, *D*, A_i and B are matrices of coefficients, and p is the length of lags of a VAR model. By multiplying with matrix D^{-1} equation (11) can be transformed into the following reduced form

$$Y_t = K + \sum_{i=1}^{p} P_i Y_{t-i} + QZ_t + E_t\,, \tag{12}$$

where $K = D^{-1}C$, $P_i = D^{-1}A_i$, $Q = D^{-1}B$, and $E_t = D^{-1}U_t$ is the reduced form errors. Equation (12) of the reduced form can be estimated from our data.

Having all coefficients matrices and matrix *D*, we can decompose each endogenous series into series of individual structural shocks using historical decomposition technique. The principle of this technique comes from the fact that the reduced form of a VAR model in (12) can be expressed by its corresponding moving average representation with recursive substitutions as follows (for $p = 2$ in our case)

$$Y_t = \mu + \sum_{i=1}^{t} \Phi_i E_i + \sum_{i=1}^{t} \Psi_i Z_i\,, \tag{13}$$

where $Y_1 = K_1 + P_1 Y_0 + P_2 Y_{-1} + QZ_1 + E_1$[3]. As $E_t = D^{-1}U_t$, equation (13) can be rewritten as

$$Y_t = \mu + \sum_{i=1}^{t} \Phi_i D^{-1} U_i + \sum_{i=1}^{t} \Psi_i Z_i = \mu + \sum_{i=1}^{t} \Gamma_i U_i + \sum_{i=1}^{t} \Psi_i Z_i, \tag{14}$$

where $\Gamma_i = \Phi_i D^{-1}$. Clearly in equation (14) each endogenous variable in the model is expressed as a sum of structural shocks up to time *t*. For details on historical decomposition technique, see Luetkepohl (2005).

3.2 Data and their pre-processing

This research is performed with the use of publically available data. The main sources of data are the databases of the Czech Statistical Office and the Czech National Bank. The exceptions are the real GDP series of Germany from Destatis database, the historical series on oil price (Europe Brent Spot Price FOB) from US Energy Information Administration database and the series on EURIBOR 12 months from website www.emmi-benchmarks.eu/euribor-org/euribor-rates.html, and finally series on CPI in Euro Area 19 and unit labor costs series for Czech Republic are from the OECD database. All series are quarterly data from 2000Q1 to 2016Q4. Each of them consists of 68 observations. All series are listed in Table 1.

For model estimation series repo rate REPO, unemployment rate UNE, unit labor costs ULC, and interbank interest rates PRIBOR and EURIBOR are kept unchanged. On the other hand, the oil price series in USD is converted into series in CZK using exchange rate USD/CZK, nominal client loans and exchange rate EUR/CZK are transformed into real quantities using price index CPI and price indices CPI and CPIEA19 respectively. After that these series and the rest are converted into corresponding year-to-year percentage changes series. All series thus have the same measure and the number of observation is reduced to 64.

[3] For other terms Φ_i is a function of P_1 and P_2, Ψ_i is a function of P_1, P_2 and *Q*.

	Notation	Full name	Unit
1	REPO	Monetary policy Repo rate	%
2	YD	Disposable income	Mil. CZK *
3	CPI	Consumer Price Index	Index
4	CONS	Households consumption	Mil CZK *
5	INV	Investments	Mil. CZK *
6	UNE	Unemployment rate	%
7	NX	Net export	Mil. CZK *
8	CRE	Client loans	Mil. CZK
9	PRIBOR	Interbank pribor one year rate	%
10	EURER	Average EURCZK exchange rate	CZK
11	ULC	Unit labor costs,year-to-year change	%
12	GDPGE	German GDP	index
13	EURIBOR	Interbank euribor one year rate	%
14	OILP	Average oil price Crude Brent Europe	USD
15	USDER	Average USDCZK exchange rate	CZK
16	CPIEA19	Consumer Price Index in Euro Area 19	index

Note: (*) denotes real variables

Table 1 List of primary data used for econometric analysis

3.3 Estimation results

We use the selected data described above to estimate a reduced VAR model with two lags. The length of lags is chosen with respect to the data available and information criterion using Bayesian technique with Minnesota prior (for more information on priors, see Canova (2007)). After that we used sign restriction approach to obtain the structural form. With these two sets of coefficient, the historical decomposition has been performed. All calculations have been executed in Matlab and the estimation as well as the decomposition are computationally very extensive. The results are displayed in Figures 1 and 2.

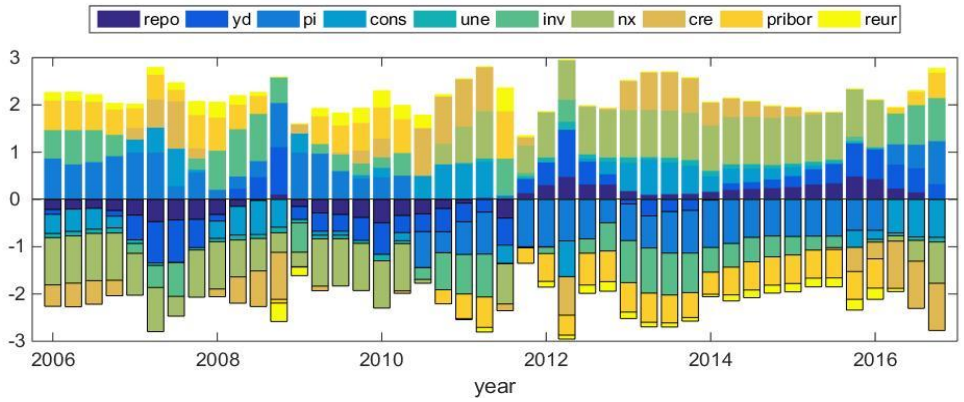


Figure 1 The contribution of endogenous shocks to PRIBOR

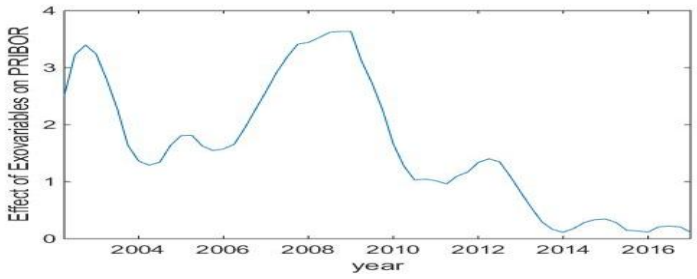


Figure 2 The contribution of exogenous variables to PRIBOR

As far as the impact of endogenous variables included in the model is concerned, in Figure 1 we can see two groups of factors which affect the money market interest rate in both opposite directions. Further, the structure of each group is not constant over time and so does the impact of each endogenous variable in the model. The positive impact of endogenous variables on money market rates are shown in case of inflation, investments and partly also

bank loans before first negative effects of crises came in 2008 and 2009. Negative shocks in pre-crisis period are represented by the dynamics of net export and consumption. The global financial crisis produced high external shock, deep decline in inflation with limited deflation expectations and pessimistic investments expectations that stay behind negative impact of net export, inflation, investment and banking loans on money market interest rates. Since the mid of 2010 net export and banking loans start to recover but inflation and investments remain weak. The application of exchange rate commitment in 2013 initiated positive impact of disposable income and unemployment (maintain positive effects of net export). In Figure 2 we can observe a positive, but declining impact of exogenous variables on the money market interest rate except the period 2006 – 2010. It reflects a pre-crisis boom in foreign demand for the gross export together with high growth of oil price followed by their declining trend after 2008. In the last several years their influence is almost negligible due to overall stabilization of these factors.

4 Conclusion

At the moment, as monetary policy interest rate as well as money market interest rates variability are found in the proximity of the zero lower bound, it raises a point on the potential factors which keep them that low as well as what may influence their apparent involatility. While money market interest rates are close to a near zero level, the optimal main monetary policy rate may exhibit significant volatility as the economy faces deep economic and/or financial changes and uncertainties of various kinds. As an attempt to give an answer to this question, the money market interest rates in the Czech economy are modeled in the framework of with a VAR model with exogenous variables. The model is estimated by Bayesian method and its structural form is obtained through sign restriction. The model's parameters are then used to decompose money market interest rate series into a series of cumulative structural shocks of each endogenous component in the model. The results show that global financial crisis and the Czech crown exchange rate commitment are two main sources of endogenous variables' shocks on money market interest rates. These endogenous variables shocks confirm high importance of net export and investment as extremely sensitive economic variables for the dynamics of the Czech open economy as well as reaching the inflation target. Long-term inflation target undershooting with deflation expectations are also the reason why the Czech National Bank inflation targeting was modified to use exchange rate commitment since 2013 to speed up demand and supply side of inflation close to inflation target.

Acknowledgements

This paper is supported by grant No. IP 100040 of the University of Economics, Prague.

References

[1] Bartolini, L., Bertola, G., Prati, A.: Day-To-Day Monetary Policy and the Volatility of the Federal Funds Interest Rate. *Journal of Money, Credit, and Banking* 34(1), 2002, 137-159.
[2] Blake, A., Mumtaz, H.: Applied Bayesian Econometrics for Central Bankers, CCBS Technical Handbook ed. 4, Bank of England, 2012.
[3] Canova, F.: *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton, 2007.
[4] Canova, F., De Nicolo, G. (2002): Monetary Disturbances Matter for Business Fluctuations in the G-7. *Journal of Monetary Economics*, 49, 1131-1159.
[5] Favero, C. A.; Rovelli, R. 2000. Modeling and Identifying Central Bank Preferences [Working Pa-per No. 148]. Bocconi: IGIER Universita Bocconi, 2000.
[6] Faust, J.: The Robustness of Identified VAR Conclusions about Money. *Carnegie Rochester Conference Series on Public Policy* 49, 207-244.
[7] Goodman, L. A. 1960. On the Exact Variance of Products. Journal of the American Statistical Association. 1960, 55 (292), 708-713.
[8] Kadiyala, R. K., Karlson, S.: Numerical methods for Estimation and Inference in Bayesian VAR Models. *Journal of Applied Econometrics* 12(2), 1997, 99 – 132.
[9] Luetkepohl, H.: *New Introduction to Multiple Time Series Analysis*, Springer - Verlag, Berlin, 2005.
[10] Rudebusch, G. D. 2002. Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia. Journal of Monetary Economics. 2002, 49(6), 1161-1187.
[11] Sack, B.: Does the Fed Act Gradually? A VAR Analysis. *Journal of Monetary Economics*. vol. 46 (1), 2000, 229-256.
[12] Srour, G. 1999. Inflation Targeting under Uncertainty [Technical Report No. 85]. Ottawa: Bank of Canada. 1999.
[13] Uhlig, H.: What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. *Journal of Monetary Economics*, 52, 381-419.
[14] Wieland, V.: Monetary Policy, Parameter Uncertainty and Optimal Learning. [*Finance and Economic Discussion* Series No. 1999-48]. Washington: Board of Governors of the Federal Reserve System, 1999.

Wavelet Method for Pricing Options with Stochastic Volatility

Dana Černá [1]

Abstract. We use the Heston stochastic volatility model for calculating the theoretical price of an option. While the Black-Scholes model assumes that the volatility of the asset is constant or a deterministic function, the Heston model assumes that the volatility is a random process. The explicit solution for the Heston model is unavailable for many types of options and therefore numerical methods have been proposed for pricing options under Heston model, e.g. Monte Carlo method, the finite difference method, the finite element method or the discontinuous Galerkin method.
The Heston model is represented by a parabolic equation. For its efficient numerical solution, we use the theta scheme for the time discretization and we propose an adaptive wavelet method for the discretization of the equation on the given time level. We construct a piecewise linear wavelet basis and use it in the scheme. The advantage of wavelets is their compression property. It means that the representation of the solution in a wavelet basis requires a small number of coefficients and the computation of the solution can be performed with the small number of parameters. Numerical example is presented for the European put option.

Keywords: Heston model, stochastic volatility, European option, wavelets, adaptive method.

JEL classification: C63, G13
AMS classification: 35K20, 65M99, 65T60

1 Introduction

Wavelet methods have been already succesfully used for solving option pricing problems, see e.g. [8, 10, 11, 15]. We also proposed and implemented a wavelet method for option pricing in our previous papers [4, 5, 6, 7]. We used the Black-Scholes model and tested the performance of the wavelet method with respect to the choice of a wavelet basis, compared an isotropic and an anisotropic approach, studied the convergence rate of the method and proposed a construction of a wavelet basis such that the Black-Scholes operator represented in this basis is sparse.

In this paper, we focus on more general Heston stochastic volatility model for calculating the price of options. Since the explicit solution of the Heston equation is known only for some special cases, it is necessary to solve it using numerical methods. Methods based on wavelets were already used for solving the Heston equation [10]. We use a different approach and propose an adaptive wavelet method that is a modification of the method from [1]. The quantitative properties of any wavelet method crucially depend on the choice of a wavelet basis. Therefore, a construction of wavelet bases is still actual [7, 16, 17] . In this paper, we adapt the piecewise linear wavelet basis with two vanishing moments from [2, 14] to the rectangle and boundary conditions representing the Heston model and use it in the scheme.

2 Heston stochastic volatility model

The Heston model assumes that the volatility is a random process and that the market price $U(S,v,t)$ of the option can be computed as the solution of the equation [9]:

$$\frac{\partial U}{\partial t}(S,v,t) - \mathcal{L}_H\left(U(S,v,t)\right) = 0, \quad S,v>0, \quad t\in(0,T), \tag{1}$$

where the operator $\mathcal{L}_H$ is given by

$$\mathcal{L}_H(U) = \frac{S^2 v}{2}\frac{\partial^2 U}{\partial S^2} + \rho\sigma S v \frac{\partial^2 U}{\partial S \partial v} + \frac{\sigma^2 v}{2}\frac{\partial^2 U}{\partial v^2} + rS\frac{\partial U}{\partial S} + \kappa\left(\tilde{\theta}-v\right)\frac{\partial U}{\partial v} - rU. \tag{2}$$

The variable S is an asset price, v is a variance of S and t is time to maturity. The parameter κ represents mean reversion, $\tilde{\theta}$ is the long term variance and σ is a volatility of the volatility, r is a risk-free rate and ρ is the correlation factor. For more details see [9, 13]. The initial and boundary conditions are of the form

$$U(S,v,0) = U_0(S,v), \quad U(0,v,t) = g(v,t), \quad \lim_{S\to\infty}\frac{\partial U}{\partial n}(S,v,t) = h_1, \quad \lim_{v\to\infty}\frac{\partial U}{\partial n}(S,v,t) = h_2, \tag{3}$$

[1]Department of Mathematics and Didactics of Mathematics, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic, dana.cerna@tul.cz

where U_0, g, h_1, h_2 are known functions and $\frac{\partial U}{\partial n}$ denotes the normal derivative.

For example, the initial condition for a European put option is given by $U(S,v,0) = \max(0, K-S)$, where K is the strike, and the boundary conditions are given by:

$$U(0,v,t) = Ke^{-rt}, \quad \lim_{S\to\infty} \frac{\partial U}{\partial S}(S,v,t) = 0, \quad \lim_{v\to\infty} \frac{\partial U}{\partial v}(S,v,t) = 0, \tag{4}$$

if Feller condition $2\kappa\theta > \sigma^2$ is satisfied [12]. We choose maximal values S^{max} and v^{max} large enough and approximate the unbounded domain $[0,\infty)\times[0,\infty)$ by a domain $\Omega = (0, S^{max})\times(0, v^{max})$. We denote the parts of the boundary of the domain Ω by

$$\begin{aligned} \Gamma_1 &= \{[S,0], S\in(0,S^{max})\}, \quad \Gamma_2 = \{[S^{max},v], v\in(0,v^{max})\}, \\ \Gamma_3 &= \{[S,v^{max}], S\in(0,S^{max})\}, \quad \Gamma_4 = \{[0,v], v\in(0,v^{max})\}, \end{aligned}$$

and we replace the initial and boundary conditions (3) with

$$U(S,v,0) = U_0(S,v), \quad U(0,v,t) = g(v,t), \quad \frac{\partial U}{\partial S}(S^{max},v,t) = h_1, \quad \frac{\partial U}{\partial v}(S,v^{max},t) = h_2. \tag{5}$$

We transform the equation (1) to the equation with homogeneous Dirichlet boundary conditions on Γ_4. Let $\tilde{U} = U - W$, where U is the solution of the equation (1) satisfying the initial and boundary conditions defined above and W is a function satisfying boundary conditions on Γ_4, e.g. $W = Ke^{-rt}$ in the case of a European put option. Let $V = \left\{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_4\right\}$. Then $\tilde{U}\in V$ and $\tilde{U}$ is the solution of the equation

$$\frac{\partial \tilde{U}}{\partial t} - \mathcal{L}_H\left(\tilde{U}\right) = f(W), \quad f(W) = -\frac{\partial W}{\partial t} + \mathcal{L}_H(W) \tag{6}$$

satisfying the initial condition $\tilde{U}(S,v,0) = U(S,v,0) - W(S,v,0)$ and homogeneous Dirichlet boundary conditions on Γ_4.

3 Semidiscretization in time and variational formulation

For discretization in time we use the θ–scheme. Let $M\in\mathbb{N}$, $\tau = T/M$, $t_l = l\tau$, $l = 0,\ldots,M$, and denote $\tilde{U}_l(S,v) = \tilde{U}(S,v,t_l)$ and $f_l(S_1,S_2) = f(W(S_1,S_2,t_l))$. The θ–scheme has the form:

$$\frac{\tilde{U}_{l+1}-\tilde{U}_l}{\tau} - \theta\mathcal{L}_H\left(\tilde{U}_{l+1}\right) - (1-\theta)\mathcal{L}_H\left(\tilde{U}_l\right) = \theta f_{l+1} + (1-\theta) f_l, \tag{7}$$

where $\theta\in[0,1]$ and $l = 0,\ldots,M-1$.

We define a bilinear form $a(u,v) = (\mathcal{L}_H(u), v)$, $u,v\in V$, and denote the standard L^2–inner product by $(\cdot,\cdot)$. The variational formulation of (7) has the form: Find $\tilde{U}_{l+1}\in V$ such that

$$\frac{\left(\tilde{U}_{l+1},v\right)}{\tau} - \theta a\left(\tilde{U}_{l+1},v\right) - (1-\theta)a\left(\tilde{U}_l,v\right) = \frac{\left(\tilde{U}_l,v\right)}{\tau} + \theta(f_{l+1},v) + (1-\theta)(f_l,v), \quad \text{for all } v\in V. \tag{8}$$

4 Construction of a piecewise linear wavelet basis

In this section, we shortly introduce the concept of a wavelet basis and propose a construction of a piecewise linear wavelet basis of the space V. The wavelet basis has two vanishing moments and is derived from wavelet bases on the interval from [2, 14].

Let H be a subspace of some Sobolev space or the L^2–space equipped with the norm $\|\cdot\|_H$, $\mathcal{J}$ be an index set such that each index $\lambda\in\mathcal{J}$ takes the form $\lambda = (j,k)$ and $|\lambda| := j$ denotes the level. A wavelet basis of the space H is defined as a family $\Psi = \{\psi_\lambda, \lambda\in\mathcal{J}\}$ such that

$i)$ Ψ is a Riesz basis for H, i.e. the closure of the span of Ψ is H and there exist constants $c, C\in(0,\infty)$ such that

$$c\|\mathbf{b}\|_2 \le \left\|\sum_{\lambda\in\mathcal{J}} b_\lambda\psi_\lambda\right\|_H \le C\|\mathbf{b}\|_2, \tag{9}$$

for all $\mathbf{b} = \{b_\lambda\}_{\lambda\in\mathcal{J}}$ such that $\|\mathbf{b}\|_2^2 = \sum_{\lambda\in\mathcal{J}} b_\lambda^2 < \infty$.

$ii)$ The functions are local in the sense that diam supp $\psi_\lambda \leq C2^{-|\lambda|}$ for all $\lambda \in \mathcal{J}$.

A wavelet basis on the interval I has typically a hierarchical structure. It consists of functions on some coarsest level j_0 called scaling functions and functions on the levels $j \geq j_0$ called wavelets. Wavelets ψ_λ, $\lambda = (j,k)$, in the inner part of the interval are typically translations and dilations of a function ψ also called wavelet, i.e. $\psi_{j,k}(x) = 2^{j/2}\psi\left(2^j x - k\right)$, and the functions near the boundary are derived from functions called boundary wavelets. In adaptive methods it is required that wavelets have vanishing moments. It means that

$$\int_I x^k \psi_{j,k}(x) = 0, \quad k = 0, \dots L-1, \tag{10}$$

where $L \geq 1$ is dependent on the type of a wavelet. A wavelet basis Ψ on the rectangle can be constructed by a tensor product of wavelet bases on intervals.

First, we construct wavelet bases for the spaces $V_1 = \left\{v \in L^2(0,1) : v(0) = 0\right\}$ and $V_2 = L^2(0,1)$ as in [2, 14]. We define scaling functions as linear B-splines. Let ϕ and ϕ_b be defined by:

$$\phi(x) = \begin{cases} x, & x \in [0,1], \\ 2-x, & x \in [1,2], \\ 0, & \text{otherwise}, \end{cases} \qquad \phi_b(x) = \begin{cases} 1-x, & x \in [0,1], \\ 0, & \text{otherwise}. \end{cases} \tag{11}$$

For $j \geq 2$ and $x \in [0,1]$ we set

$$\begin{aligned} \phi^1_{j,k}(x) &= 2^{j/2}\phi(2^j x - k + 1),\ k = 1, ..., 2^j - 1, \\ \phi^1_{j,2^j}(x) &= 2^{j/2}\phi_b\left(2^j(1-x)\right), \\ \phi^2_{j,1}(x) &= 2^{j/2}\phi_b(2^j x), \\ \phi^2_{j,k}(x) &= 2^{j/2}\phi(2^j x - k + 2),\ k = 2, ..., 2^j, \\ \phi^2_{j,2^j+1}(x) &= 2^{j/2}\phi_b\left(2^j(1-x)\right), \end{aligned} \tag{12}$$

and

$$\Phi^1_j = \left\{\phi^1_{j,k}, k = 1, \dots, 2^j\right\}, \quad \Phi^2_j = \left\{\phi^2_{j,k}, k = 1, \dots, 2^j + 1\right\}. \tag{13}$$

We define a wavelet ψ and a boundary wavelets ψ_{b1} and ψ_{b2} as

$$\begin{aligned} \psi(x) &= -\frac{1}{4}\phi(2x) - \frac{1}{2}\phi(2x-1) + \frac{3}{2}\phi(2x-2) - \frac{1}{2}\phi(2x-3) - \frac{1}{4}\phi(2x-4), \\ \psi_{b1}(x) &= \frac{3}{2}\phi_b(2x) - \frac{9}{8}\phi(2x) + \frac{1}{4}\phi(2x-1) + \frac{1}{8}\phi(2x-2), \\ \psi_{b2}(x) &= \frac{3}{2}\phi(2x) - \phi(2x-1) - \frac{1}{2}\phi(2x-2), \end{aligned} \tag{14}$$

Then supp $\psi = [0,3]$, supp $\psi_{b1} =$ supp $\psi_{b2} = [0,2]$, and these wavelets have two vanishing moments. For $j \geq 2$, $i = 1, 2$, and $x \in [0,1]$ we define

$$\begin{aligned} \psi^i_{j,1}(x) &= 2^{j/2}\psi_{bi}(2^j x), \\ \psi^i_{j,k}(x) &= 2^{j/2}\psi(2^j x - k + 2), k = 2, ..., 2^j - 1, \\ \psi^i_{j,2^j}(x) &= 2^{j/2}\psi_{b1}(2^j(1-x)), \end{aligned} \tag{15}$$

and

$$\Psi^i_j = \left\{\psi^i_{j,k}, k = 1, \dots, 2^j\right\}. \tag{16}$$

Hence, Ψ^1_j and Ψ^2_j differs only in the left boundary wavelet. The inner wavelets and the right boundary wavelets are the same. The graphs of the wavelets $\psi^1_{j,k}$ and $\psi^2_{j,k}$ on the level $j = 2$ are displayed in Figure 1. It was proved in [2, 14] that the set

$$\Psi^i = \Phi^i_2 \cup \bigcup_{j=2}^{\infty} \Psi^i_j \tag{17}$$

is a wavelet basis of the space V_i.

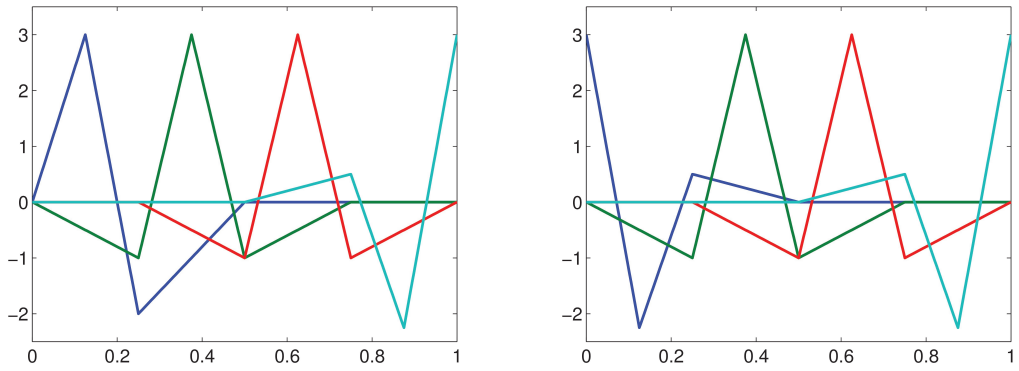


Figure 1 Wavelets $\psi^1_{j,k}$ (left) and $\psi^2_{j,k}$ (right) on the level $j = 2$.

Now, using a simple linear transformation and the tensor product, we obtain a wavelet basis of the space V. For $S \in [0, S^{max}]$ and $v \in [0, v^{max}]$ let us define

$$\begin{aligned} \phi^S_{j,k}(S) &= \phi^1_{j,k}\left(\frac{S}{S^{max}}\right), \quad \psi^S_{j,k}(S) = \psi^1_{j,k}\left(\frac{S}{S^{max}}\right), \\ \phi^v_{j,k}(v) &= \phi^2_{j,k}\left(\frac{v}{v^{max}}\right), \quad \psi^v_{j,k}(v) = \psi^2_{j,k}\left(\frac{v}{v^{max}}\right). \end{aligned} \tag{18}$$

Then the set

$$\begin{aligned} \Psi &= \left\{\phi^S_{2,k}(S)\cdot\phi^v_{2,l}(v), k \in I^1_2, l \in I^2_2\right\} \cup \left\{\phi^S_{2,k}(S)\cdot\psi^v_{j,l}(v), k \in I^1_2, j \geq 2, l \in J_j\right\} \\ &\cup \left\{\psi^S_{j,k}(S)\cdot\phi^v_{2,l}(v), j \geq 2, k \in J_j, l \in I^2_2, \right\} \cup \left\{\psi^S_{i,k}(S)\cdot\psi^v_{j,l}(v), i, j \geq 2, k \in J_i, l \in J_j\right\}, \end{aligned} \tag{19}$$

$I^1_2 = \{1, \ldots, 4\}$, $I^2_2 = \{1, \ldots, 5\}$, $J_j = \left\{1, \ldots, 2^j\right\}$, is a wavelet basis for the space V equipped with the L^2-norm and Ψ, when normalized with respect to the H^1-norm, is a wavelet basis for the space V equipped with the H^1-norm.

5 Adaptive wavelet method

We use an adaptive wavelet method for discretization of the equation (7) that is a modification of the method from [1]. This method is different from classical adaptive methods that are based on mesh refinement according to error estimates. We start with a variational formulation but instead of turning to a finite dimensional approximation, we transform the continuous problem into an infinite-dimensional problem using the suitable wavelet basis. Then we propose an iteration scheme for this problem. Finally, we replace all infinite-dimensional quantities by finitely supported ones and we use the routine for an application of an infinite matrix approximately.

Let $\Psi = \{\psi_\lambda, \lambda \in \mathcal{J}\}$ be a wavelet basis from the previous section and $\mathbf{u} = \{u_\lambda\}_{\lambda\in\mathcal{J}}$ be the coefficients of the solution $\tilde{U}_{l+1}$ of the problem (8) in a basis Ψ, i.e.

$$\tilde{U}_{l+1} = \sum_{\lambda\in\mathcal{J}} u_\lambda \psi_\lambda. \tag{20}$$

Using (8), (20), and setting $v = \psi_\mu$ we obtain the infinite matrix equation $\mathbf{A}\mathbf{u} = \mathbf{f}$ with

$$\mathbf{A}_{\mu,\lambda} = \frac{(\psi_\lambda, \psi_\mu)}{\tau} - \theta a(\psi_\lambda, \psi_\mu), \quad \mu, \lambda \in \mathcal{J}, \tag{21}$$

and

$$\mathbf{f}_\mu = (1-\theta)\, a\left(\tilde{U}_l, \psi_\mu\right) + \frac{\left(\tilde{U}_l, \psi_\mu\right)}{\tau} + \theta(f_{l+1}, \psi_\mu) + (1-\theta)(f_l, \psi_\mu), \quad \mu, \lambda \in \mathcal{J}. \tag{22}$$

It is clear that $\mathbf{f}$ and $\mathbf{u}$ depend on the time level t_l, but we omit the index l to simplify notation. For solving this infinite-dimensional problem we use the method of generalized residuals (GMRES). We numerically tested several iteration methods and the GMRES seems to be the most efficient. We use the Jacobi diagonal preconditioner $\mathbf{D}$, where the diagonal elements of $\mathbf{D}$ satisfy $\mathbf{D}_{\lambda,\lambda} = \sqrt{\mathbf{A}_{\lambda,\lambda}}$. We obtain the preconditioned system

$$\tilde{\mathbf{A}}\tilde{\mathbf{u}} = \tilde{\mathbf{f}} \tag{23}$$

with

$$\tilde{\mathbf{A}} = \mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}, \quad \tilde{\mathbf{f}} = \mathbf{D}^{-1}\mathbf{f}, \quad \tilde{\mathbf{u}} = \mathbf{D}\mathbf{u}. \tag{24}$$

For the given time level t_l the algorithm for solving the infinite-dimensional problem comprises the following steps:

1. Compute sparse representation $\tilde{\mathbf{f}}_j$ of the right-hand side $\tilde{\mathbf{f}}$ with the error smaller than given tolerance ϵ_j.

2. Compute several steps of GMRES iterations for solving the system $\tilde{\mathbf{A}}\mathbf{v} = \tilde{\mathbf{f}}_j$ with the initial vector $\mathbf{v}_j$. Each iteration of GMRES requires multiplication of the infinite-dimensional matrix with a vector. It is computed approximately with the given tolerance $\hat{\epsilon}_j$ by the method from [3]. We denote the resulting vector by $\mathbf{z}$.

3. Compute sparse representation $\mathbf{v}_{j+1}$ of $\mathbf{z}$ with the error smaller than $\tilde{\epsilon}_j$.

We repeat the steps 1., 2., and 3. until the residual is not smaller than the required error. Since we work with the sparse representation of the right-hand side and the sparse representation of the vector representing the solution, the method is adaptive. The computation of a sparse representation is simple and it insists in thresholding the smallest coefficients and working only with the largest coefficients. It is known that the coefficients in the wavelet basis are small in regions where the function is smooth and large in regions where the function has some singularity.

6 Numerical example

We use the proposed scheme for computing the price of a European put option with the same parameters as in [9, 12], i.e. the mean reversion $\kappa = 2$, long run variance $\tilde{\theta} = 0.01$, current variance $v(t) = 0.01$, correlation $\rho = 0.5$, volatility of volatility $\sigma = 0.1$, option maturity 0.5 year, interest rate $r = 0$ and the strike price $K = 100$.

We choose $S^{max} = 400$, $v^{max} = 1$, $\theta = 0.5$ and the time step $\tau = 1/3650$. We use the linear spline wavelet basis proposed in Section 4. The solution U at time $t = 0$, i.e. the function representing the initial condition, and the solution U for $t = 0.5$ are displayed in Figure 2.

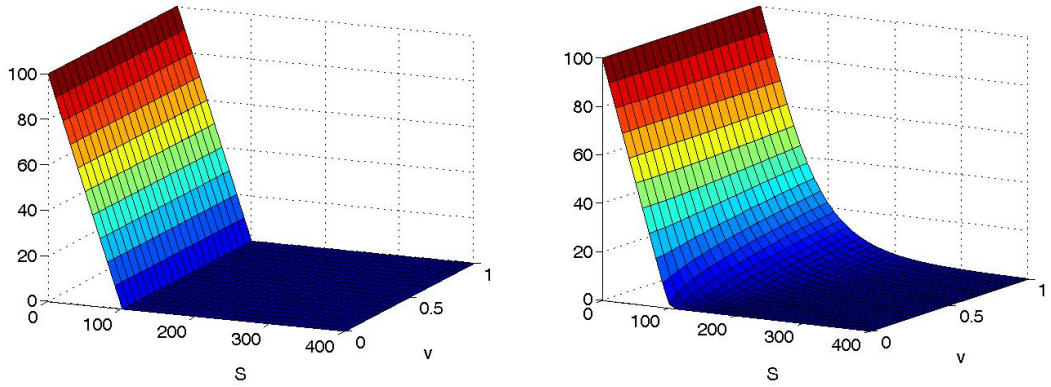


Figure 2 The plot of the value of the option $U(S, v, t)$ for $t = 0$ (left) and $t = 0.5$ (right).

It can be seen that the function $U(S, v, 0)$ has not derivative on the line $\{100\} \times (0, v_{max})$. Therefore the largest wavelet coefficients correspond to wavelets with supports in regions near this line and wavelet coefficients are small for wavelets that are not located in these regions. Thus many wavelet coefficients are thresholded and the representation of the solution is sparse. The number of parameters representing the solution U in Figure 2 is 617. In Table 1 the values of U are listed for $t = 1/2$, $v = 0.01$ and several values of S and ρ and for comparison, the values for the Black-Scholes model with the corresponding volatility $\sigma = 0.1$ are listed.

	Heston				Black-
$S\backslash\rho$	-0.9	-0.5	0.5	0.9	Scholes
90	10.058	10.128	10.293	10.355	10.201
100	2.778	2.784	2.797	2.801	2.820
110	0.475	0.407	0.213	0.122	0.305

Table 1 The solution $U(S, 0.01, 0.5)$ for several values of S and ρ and the values for the Black-Scholes model.

7 Conclusion

We extended our research to option pricing under the Heston stochastic volatility model. We constructed a piecewise linear wavelet basis with two vanishing moments on the rectangle $(0, S_{max}) \times (0, v_{max})$ that is adapted to boundary conditions of the type (4). We proposed and implemented an adaptive wavelet method with this basis for a numerical solution of the Heston equation. We presented numerical example for the European put option and compared the results with the Black-Scholes model. The main advantage of the method is a small number of parameters representing the solution with desired accuracy. Our future aim is to develop and analyze the efficient solver for multi-asset options with stochastic volatility and for other option pricing models such as the Hull-White model, the Stein and Stein model and the Lévy model and to compare the results.

Acknowledgements

This work was supported by grant No. GA16-09541S of the Czech Science Foundation.

References

[1] Cohen, A., Dahmen, W., and DeVore, R.: Adaptive wavelet methods II - beyond the elliptic case, *Found. Math.* **2** (2002), 203–245.

[2] Černá, D., and Finěk, V: Construction of optimally conditioned cubic spline wavelets on the interval. *Adv. Comput. Math.* **34** (2011), 219–252.

[3] Černá, D., and Finěk, V.: Approximate multiplication in adaptive wavelet methods. *Cent. Eur. J. Math.* **11** (2013), 972–983.

[4] Černá, D.: Numerical solution of the Black-Scholes equation using cubic spline wavelets. In *AIP Conference Proceedings* **1789**; (V. Pasheva, N. Popivanov, and G. Venkov, eds.), American Institute of Physics, New York, 2016, No. 030001 (8 pages).

[5] Černá, D., and Finěk, V.: Adaptive wavelet method for the Black-Scholes equation of European options. In *Proceedings of the 34th International Conference on Mathematical Methods in Economics*, (A. Kocourek, and M. Vavroušek), Liberec, 2016, 120–125.

[6] Černá, D.: Wavelet methods for pricing European options. In *Proceedings of the 8th International Scientific Conference Managing and Modelling of Financial Risks*, (M. Čulík), Ostrava, 2017, 129–136.

[7] Černá, D., and Finěk, V.: Sparse wavelet representation of differential operators with piecewise polynomial coefficients. *Axioms* **6** (2017), No. 4 (21 pages).

[8] Finěk, V.: Fractional step method for wavelet based solution of Black-Scholes equation. In *AIP Conference Proceedings* **1789**; (V. Pasheva, N. Popivanov, and G. Venkov, eds.), American Institute of Physics, New York, 2016, No. 030007 (4 pages).

[9] Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6** (1993), 327–343.

[10] Hilber, N.: Stabilized wavelet methods for option pricing in high dimensional stochastic volatility models. Ph.D. thesis, ETH Zurich, 2009.

[11] Hilber, N., Reichmann, O., Schwab, C., and Winter, C.: *Computational Methods for Quantitative Finance*. Springer, Berlin, 2013.

[12] Hozman, J., and Tichý, T.: A discontinuous Galerkin method for numerical pricing of European options under Heston stochastic volatility. In *AIP Conference Proceedings* **1789**; (V. Pasheva, N. Popivanov, and G. Venkov, eds.), American Institute of Physics, New York, 2016, No. 030003 (7 pages).

[13] Mikhailov, S., and Nögel, U.: Heston's stochastic volatility model implementation, calibration and some extensions. *Wilmott Magazine* (2005), 74–79.

[14] Primbs, M.: New stable biorthogonal spline-wavelets on the interval. *Result. Math.* **57** (2010), 121–162.

[15] Rometsch, M.: *A wavelet tour of option pricing*. PhD thesis, Universität Ulm, 2010.

[16] Shumilov, B.M.: Cubic multiwavelets orthogonal to polynomials and a splitting algorithm. *Numer. Anal. and Appl.* **6** (2013), 247–259.

[17] Shumilov, B.M.: Semi-orthogonal spline-wavelets with derivatives and the algorithm with splitting. *Numer. Analys. Appl.* **10** (2017), 90–100.

On the limit identification region for regression parameters in linear regression with interval-valued dependent variable

Michal Černý, Miroslav Rada, Ondřej Sokol, Vladimír Holý[1]

Abstract. We consider the linear regression model where the dependent variable cannot be observed. Instead, only a lower and upper bound (interval for the dependent variable) is observable. In this case, the vector of regression parameters may be only partially identified, or unidentified at all. We make as weak assumptions as possible on the process generating the observable intervals to make the regression parameters partially identified. First we consider the finite-sample identification region for the OLS-estimator. Then we construct the identification region for the vector of regression parameters as a limit of the finite-sample regions. We derive explicit bounds for the limit set. Although we construct the bounds only in case of two regressors, the idea can be pushed further to regression models with a general number of regressors. The derived bounds depend only on observable quantities (such as the variance of a regressor or covariance between the regressor and the widths of the intervals of the dependent varible) and can be estimated from finite samples. We also suggest some possible ways of further research to (i) refine the bounds suggested in this paper, (ii) tighten the bounds using additional assumptions with identifying power.

Keywords: linear regression; interval data; partial identification

JEL classification: C13
AMS classification: 62J86

1 Introduction

Generally, a parameter p of a distribution is *partially identified* if observable data do not allow us to estimate p consistently, but we are able to retrieve at least some 'nontrivial' information about p. An example is when we can consistently estimate only a bound on p, or when we can consistently estimate a nontrivial set P such that $p \in P$. (Such set P is often called *identification region* for p.)

The partial identification problem often arises when we work with a joint distribution (X, Y) of a pair of (possibly vector-valued) variables, where one of them is observable and the other one is not. Say, for example, that X is observable and Y is not. The task is to make inference about the distribution of Y. Here, *observability* means that one can make a corresponding random sample $x_1, \ldots, x_n$ from the distribution of X and all estimators and test statistics are allowed to be functions of $x_1, \ldots, x_n$ only.

This general setup involves many cases. For example, if X, Y are independent, then information about X does not reveal any information about Y. On the contrary, if Y and X are strongly correlated (or linked otherwise), then inference about X might give also a lot of information about Y.

Say that we are interested in a parameter $p \equiv p(Y)$ of the distribution of Y (such as mean, a quantile, variance or another characteristic), but one can construct estimators of p only as a function of $x_1, \ldots, x_n$. Then it might happen that:
(a) p is fully identified, meaning that there exists an estimator $\widehat{p}(x_1, \ldots, x_n)$ of p, which is consistent;
(b) p is partially identified. The partial identification corresponds to the situation that the sample $x_1, \ldots, x_n$ reveals some nontrivial information about p, but p is not consistently (pointwise) estimable.
(c) Finally it might happen that p is unidentified at all, meaning that the sample $x_1, \ldots, x_n$ does not allow us to say anything nontrivial about Y.

We refer the reader to [6, 9, 10] as good sources on partial identification.

Example 1. Consider that X measures heights of men and Y heights of women. Assume that $\mathsf{E}Y$ is known to satisfy $\mathsf{E}Y \leq \mathsf{E}X$. When we can observe only a sample $x_1, \ldots, x_n$ of men, then $\frac{1}{n}\sum_{i=1}^{n} x_i$ is a consisent estimator of an upper bound on $\mathsf{E}Y$, but $\mathsf{E}Y$ is not pointwise estimable. This is a trivial example of the case when $\mathsf{E}Y$ is partially identified.

[1]All authors: University of Economics, Prague, Faculty of Informatics & Statistics, Department of Econometrics, Winston Churchill Square 4, 13067 Prague, Czech Republic, {cernym, miroslav.rada, ondrej.sokol, vladimir.holy}@vse.cz

Example 2. This example follows the lines of [6]. Consider a survey where Y is the unobservable distribution of interest and the task is to estimate $\mathsf{E}Y$. Consider also a 0-1 variable R such that we can observe $X = (X^1 := R, X^2 := RY)$. For example, let Y measure election preferences across the entire population. Assume that R is an indicator whether a questioned person in the survey does $(R = 1)$ or does not $(R = 0)$ respond. We can sample at random from X, getting a sample $x_i = (r_i, r_i y_i)_{i=1,\dots,n}$. In general, the conditional distributions $[Y|R = 0]$ and $[Y|R = 1]$ can be different. But we can write

$$\mathsf{E}Y = \mathsf{E}[Y|R = 0] \cdot (1 - \mathsf{E}R) + \mathsf{E}[Y|R = 1] \cdot \mathsf{E}R.$$

Turning back to the finite-sample case, let $I^1 := \{i \in \{1, \dots, n\} \mid r_i = 1\}$. Taking $\widehat{p} = \sum_{i=1}^n r_i$ as a consistent estimator of $p := \Pr[R = 1] = \mathsf{E}R$ (the 'response rate'), we have (under some mild assumptions, which are not essential in this introductory section)

$$\mathsf{E}Y = \lim_{n\to\infty} \left\{ \mathsf{E}[Y|R = 0](1 - \widehat{p}) + \left(\frac{1}{|I^1|} \sum_{i\in I^1} r_i y_i \right) \widehat{p} \right\} = \left\{ \mathsf{E}[Y|R = 0](1 - \widehat{p}) + \left(\frac{1}{|I^1|} \sum_{i\in I^1} y_i \right) \widehat{p} \right\},$$

where the limit is understood in probability. (For the sake of brevity, the 'in probability' meaning of limits will not be repeated in the sequel, but tacitly the limits should be understood in this way.) Assume further that $\mathsf{E}[Y|R = 0]$ can be bounded apriorily by a pair of known constants $\underline{y}, \overline{y}$ (for example, if Y is a 0-1 variable, then we can clearly put $\underline{y} = 0, \overline{y} = 1$). We get a pair of bounds

$$\mathsf{E}Y \le \lim_{n\to\infty} \left\{ \overline{y}(1 - \widehat{p}) + \left(\frac{1}{|I^1|} \sum_{i\in I^1} y_i \right) \widehat{p} \right\},$$

$$\mathsf{E}Y \ge \lim_{n\to\infty} \left\{ \underline{y}(1 - \widehat{p}) + \left(\frac{1}{|I^1|} \sum_{i\in I^1} y_i \right) \widehat{p} \right\}.$$

Thus, the parameter of interest $\mathsf{E}Y$ is partially identified if $p > 0$ since we have constructed a pair of consistently estimable nontrivial bounds for it; and is fully identified if $p = 1$ since the bounds are equal in that case.

Remark. It is obvious that identifiability or partial identifiability depends on what assumptions we are able to make about the unobservable quantities. For example, if we could make the "missing-at-random" assumption, meaning that the conditional distributions $[Y|R = 0]$ and $[Y|R = 1]$ are the same, then we could simply drop the non-responses and get $\mathsf{E}Y = \mathsf{E}[Y|R = 1] = \lim_{n\to\infty}(1/|I^1|) \sum_{i\in I^1} y_i$. The parameter $\mathsf{E}Y$ would be pointwise identified whenever $p > 0$.

2 Linear regression with interval-valued outcomes

Now we turn the attention to the partial identification problem in linear regression with interval-valued endogenous variable. To fix the setup, we first consider the linear regression relationship

$$y_i = x_i^{\mathrm{T}}\theta + \varepsilon_i, \quad i = 1, \dots, n,$$

where the vector of regression parameters θ is to be estimated. We assume that the (random) regressors x_i are observable, while the dependent variable y_i is *not*; instead, we can only observe bounds $\underline{y}_i, \overline{y}_i$ such that almost surely (a.s.)

$$\underline{y}_i \le y_i \le \overline{y}_i, \quad i = 1, \dots, n. \tag{1}$$

This setup fits well in the (X, Y)-model from Section 1. We have a random process generating regressors x_i which are observable. We have a random process generating the bounds $\underline{y}_i, \overline{y}_i$, which are observable too. The rest is unobservable (namely, the y_i's are unobservable), meaning that the estimators $\widehat{\theta}$ of θ are allowed to be functions of the observable quantities $x_i, \underline{y}_i, \overline{y}_i$ only.

Now we make an additional assumption that the distribution of $(x_i, \varepsilon_i)_{i=1,\dots,n}$ is such that Ordinary Least Squares $\widehat{\theta}^{\mathrm{OLS}} \equiv \widehat{\theta}^{\mathrm{OLS}}(x_1, \dots, x_n; y_1, \dots, y_n)$ is a consistent estimator of θ. Contrary to the textbook case, $\widehat{\theta}^{\mathrm{OLS}}$ is an unobservable quantity since in our model the y_i's are unobservable. Thus, $\widehat{\theta}^{\mathrm{OLS}}$ is a 'prohibited' estimator. But what is observable is the set

$$\widehat{\Theta} := \{\widehat{\theta}^{\mathrm{OLS}}(x_1, \dots, x_n; \upsilon_1, \dots, \upsilon_n) \mid \underline{y}_i \le \upsilon_i \le \overline{y}_i,\ i = 1, \dots, n\}. \tag{2}$$

Moreover, we clearly have

$$\widehat{\theta}^{\mathrm{OLS}}(x_1, \dots, x_n; y_1, \dots, y_n) \in \widehat{\Theta} \quad \text{a.s.} \tag{3}$$

and generally $\widehat{\Theta}$ is the smallest set satisfying (3). [Indeed, if one removes a point θ^0 from $\widehat{\Theta}$, then — without additional assumptions — one cannot rule out that the distributions of x_i, y_i are concentrated in certain points x_i^0, y_i^0 such that $\widehat{\theta}^{\mathrm{OLS}}(x_1^0, \dots, x_n^0; y_1^0, \dots, y_n^0) = \theta^0$.] We can say that $\widehat{\Theta}$ is a *tight bound* for $\widehat{\theta}^{\mathrm{OLS}}$. We call it *identification region* for $\widehat{\theta}^{\mathrm{OLS}}$. (In [7], such sets are interpreted as possibilistic generalizations of real-valued statistics under interval uncertainty; see also [1].)

To emphasize that $\widehat{\Theta}$ depends on the number of observations n, we write $\widehat{\Theta}^n$. The basic question of the paper is whether there exists a limit set Θ such that $\widehat{\Theta}^n \to \Theta$ when $n \to \infty$. (This is said very loosely, since the in-probability convergence of a sequence of sets can be formalized in various ways; but later we will state our results more carefully, but in a way sufficient for the presentation of the main idea.) If this is the case and the limit set Θ is nontrivial (for example, is compact), then we have $\theta \in \Theta$ by (3) and the vector of regression parameters θ is partially identified.

Generally, this is not the case — simply imagine that the widths $L_i := \overline{y}_i - \underline{y}_i$ of the observable intervals covering y_i grow unboundedly in a way that the "limit" of $\widehat{\Theta}_n$ covers the entire parameter space. Another extreme case is when $\widehat{\Theta}_n$ converges to a singleton; then, θ would be fully identified. In the next section we study the intermediate case.

3 The case with two parameters

Here we restrict ourselves to a very simple linear regression model

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i, \quad i = 1, \dots, n, \tag{4}$$

with a single stochastic regressor. We make traditional assumptions:

(i) $x_1, \dots, x_n$ are independent and identically distributed (iid) with $\mathsf{E}x_i =: \xi$ and $\operatorname{var} x_i =: \pi^2 < \infty$,
(ii) $\varepsilon_1, \dots, \varepsilon_n$ are iid with $\mathsf{E}\varepsilon_i = 0$, $\operatorname{var} x_i =: \sigma^2 < \infty$ and $\operatorname{cov}(x_i, \varepsilon_i) = 0$.

Assumptions (i) and (ii) imply that $\widehat{\theta}^{\mathrm{OLS}} \to \theta$ and thus $\theta \in \Theta$ (if the limit set Θ of $\widehat{\Theta}^n$ exists) by (3).

Example 3. In this example we show that the limit set Θ need not exist. Also, the example illustrates what can happen in general. For the sake of simplicity of the example, assume that it is known that $\theta_2 = 0$. Then (4) reduces to the simple model of position $y_i = \theta + \varepsilon_i$. Then, the OLS-estimator of θ has the form $\widehat{\theta}^{\mathrm{OLS}} = \frac{1}{n}\sum_{i=1}^n y_i$ and the identified set has the form

$$\widehat{\Theta}^n = \left\{ \frac{1}{n}\sum_{i=1}^n \upsilon_i \;\middle|\; \underline{y}_i \le \upsilon_i \le \overline{y}_i \right\} = \left[\frac{1}{n}\sum_{i=1}^n \underline{y}_i,\ \frac{1}{n}\sum_{i=1}^n \overline{y}_i \right]. \tag{5}$$

Now $\widehat{\Theta}^n$ converges to a bounded set Θ only if both bounds $\frac{1}{n}\sum_{i=1}^n \underline{y}_i$ and $\frac{1}{n}\sum_{i=1}^n \overline{y}_i$ converge. And this need not be the case: consider, for example,

$$[\underline{y}_i, \overline{y}_i] = \begin{cases} [y_i, y_i + 1] & \text{if the number of decimal digits of } i \text{ is odd,} \\ [y_i - 1, y_i] & \text{if the number of decimal digits of } i \text{ is even.} \end{cases}$$

This (somehow artificial) example also shows that the limit set Θ need not exist even if the widths $L_i := \overline{y}_i - \underline{y}_i$ are constant.

But still, we can state at least the following property: when $\widehat{\Theta}^1, \widehat{\Theta}^2, \dots$ is the sequence (5) and whenever we choose a sequence of representatives $\{\widetilde{\theta}^i \in \widehat{\Theta}^i,\ i = 1, 2, \dots\}$, then every accumulation point θ^* of the sequence $\{\widetilde{\theta}^i\}_{i=1,\dots,n}$ satisfies

$$\theta - 1 \le \theta^* \le \theta + 1. \tag{6}$$

This is the 'more careful' formalization of the intuitive fact that if all intervals have widths

$$L_i = 1, \tag{7}$$

then the position parameter θ generally cannot be estimated consistently, but can be asymptotically estimated with error at most ± 1. In addition, this example shows that — if nothing about the interval-generating process in known except for (7) — it might be a 'good idea' to select the representatives as midpoints $\widetilde{\theta}^i = \frac{1}{2n}\sum_{i=1}^n (\underline{y}_i + \overline{y}_i)$. Now (6) can be strengthened to $\theta - \frac{1}{2} \le \theta^* \le \theta + \frac{1}{2}$. (This is basically the idea of Section 8 from [4].) This bound cannot be improved in general. *End of Example 3.*

Now we turn our attention to the main result of this paper. In addition to (i) – (ii), we also make the following assumption.

(iii) Let $L_i := \overline{y}_i - \underline{y}_i$ denote the width of the interval $[\underline{y}_i, \overline{y}_i]$. We assume that $L_1, L_2, \ldots$ are iid with $\mathsf{E}L_i =: \lambda$ and $\operatorname{var} L_i < \infty$.

We denote $\nu := \operatorname{cov}(x_i, L_i)$. Generally, we admit that the width L_i may be correlated with the regressor x_i.

We also make the following simplifying assumption:

$$x_i \geq 0 \quad \text{a.s.} \tag{8}$$

This assumption can be relaxed at the cost of some more algebra in the proof of Theorem 1. However, here we are trying to keep things as simple as possible to highlight the main idea.

As in Example 3, let $\widetilde{\theta}^i \in \widehat{\Theta}^i$ be a sequence of representatives from the identification regions for the OLS-estimator of (4) and let θ^* be an accumulation point of $\{\widetilde{\theta}^i\}_{i=1,2,\ldots}$.

Theorem 1. *The accumulation point $\theta^* = (\theta_1^*, \theta_2^*)^{\mathrm{T}}$ can be bounded as follows:*

$$\theta_1 - \xi\frac{\nu + 2\lambda\xi}{\pi^2} \leq \theta_1^* \leq \theta_1 + \xi\frac{\nu + 2\lambda\xi}{\pi^2}, \quad \theta_2 - \frac{\nu + 2\lambda\xi}{\pi^2} \leq \theta_2^* \leq \theta_2 + \frac{\nu + 2\lambda\xi}{\pi^2}. \tag{9}$$

Proof. From (2) it follows that

$$\widehat{\Theta} = \{(\widehat{\theta}_1, \widehat{\theta}_2)^{\mathrm{T}} \mid \underline{y}_i \leq y_i \leq \overline{y}_i,\ i = 1, \ldots, n\},$$

where

$$\widehat{\theta}_2 = \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n y_i\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}, \quad \widehat{\theta}_1 = \left(\frac{1}{n}\sum_{i=1}^n y_i\right) - \widehat{\theta}_2 \cdot \frac{1}{n}\sum_{i=1}^n x_i.$$

Let $y_i \in [\underline{y}_i, \overline{y}_i]$ be arbitrary. We derive an upper bound for $\widehat{\theta}_2$:

$$\begin{aligned}
\widehat{\theta}_2 &= \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n y_i\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \\
&\leq \frac{\frac{1}{n}\sum_{i=1}^n x_i \overline{y}_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n \underline{y}_i\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \\
&\leq \frac{\frac{1}{n}\sum_{i=1}^n x_i (y_i + L_i) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n (y_i - L_i)\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i + \frac{1}{n}\sum_{i=1}^n x_i L_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n y_i\right) + \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n L_i\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n y_i\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} + \frac{\frac{1}{n}\sum_{i=1}^n x_i L_i + \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n L_i\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n y_i\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \\
&\qquad + \frac{\frac{1}{n}\sum_{i=1}^n x_i L_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n L_i\right) + 2\left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n L_i\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \\
&\overset{n\to\infty}{\longrightarrow} \theta_2 + \frac{\nu + 2\xi\lambda}{\pi^2}.
\end{aligned}$$

In the limit $n \to \infty$, we have used

$$\frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n y_i\right)}{\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2} \to \theta_2$$

(this is the traditional consistency argument for OLS), $\left[\frac{1}{n}\sum_{i=1}^n x_i L_i - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{1}{n}\sum_{i=1}^n L_i\right)\right] \to \nu$ (the sample covariance between x and L converges to the true value ν), $\left[\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2\right] \to \pi^2$ (the sample variance of x converges to the true value π^2), $\frac{1}{n}\sum_{i=1}^n x_i \to \xi$ and $\frac{1}{n}\sum_{i=1}^n L_i \to \lambda$. Thus, in the limit, the accumulation point θ_2^* can be upper-bounded by $\theta_2 + (\nu + 2\xi\lambda)/\pi^2$.

The derivation of the lower bound for θ_2^* is similar.

Since $\widehat{\theta}_1$ is linear in y_i, its bounds can be derived analogously. □

Remark. In the proof, the auxiliary assumption (8) does not matter in the derivation of the bounds for θ_2^*, because an x-shift of data does not change the slope of the regression line. However, in the derivation of the bounds for θ_1^*, one has to distinguish the cases $\xi > 0$ and $\xi < 0$ (which is straightforward).

4 Discussion

What is essential is that *the bounds (9) are consistently estimable*. Indeed, the bounds are functions of ξ, π^2, ν, λ. The quantities ξ, π^2 are estimable since the regressor x is observable. The widths L_i of the intervals are observable, too; thus λ and ν are estimable as well. It follows that we can consistently estimate the bounds (9) and have an idea how much 'limit imprecision' is brought to the regression model by the interval uncertainty in the endogenous variable.

Observe also the following interesting fact — the bounds (9) *do not* depend on $\mathrm{var}\, L_i$ (we just need the variance to be finite).

It is also worth emphasizing that the bounds work even if the widths L_i are correlated with x_i and/or ε_i. Basically, the bounds (9) are wider if x_i and L_i are positively correlated. (This is not surprising.) However, what is interesting that we needed no assumptions about the joint distribution of L_i and ε_i and that the limit bounds are the same even if these quantities are correlated.

The assumptions on the behavior of x_i, ε_i are in a sense standard ones. The assumptions on $\underline{y}_i, \overline{y}_i$ are 'as weak as possible' — we just assumed (1) and (iii). This makes the model very general; the (random) process generating the intervals $[\underline{y}_i, \overline{y}_i]$ can be 'almost arbitrary'. Indeed, (1) holds true by definition of the model and (iii) is required for the Central Limit Theorem. Nothing more is needed. This is in line with the so-called *credibility paradigm* [6]; we did not impose any special assumptions on the mechanism generating the intervals $[\underline{y}_i, \overline{y}_i]$. On the one hand, the assumption on the finiteness of variance of L_i in (iii) is testable only hardly; on the other hand, nonnegative distributions with finite mean and infinite variance are quite rare and unnatural. Consider, for example, the distribution with density $\varphi(\zeta) = (3/2)\zeta^{-5/2} \cdot \mathbb{I}\{\zeta \geq 1\}$, where $\mathbb{I}\{\cdot\}$ stands for the 0-1 indicator function. In theory, the widths $L_1, L_2, \ldots$ could be sampled from this distribution; but in practice one would hardly find an example where this distribution of the widths would have a natural justification.

The only assumption, which might be considered as restrictive, is the independence of L_i, L_j with $i \neq j$. At least, this property is testable. Obviously, it is worth devoting further effort to its relaxation, at least for the case of some simple dependence structures.

In Example 3 we discussed that the sequence of sets $\widehat{\Theta}^n$ need not converge with $n \to \infty$. This was why we formulated Theorem 1 in terms of the accumulation point θ^*. However, another statement could be as follows: for every $\delta > 0$, $\Pr[\widehat{\Theta}^n \subseteq C_\delta] \to 1$ with $n \to \infty$, where

$$C_\delta = \left[\theta_1 - \xi\frac{\nu + 2\lambda\xi}{\pi^2} - \delta,\ \theta_1 + \xi\frac{\nu + 2\lambda\xi}{\pi^2} + \delta\right] \times \left[\theta_2 - \frac{\nu + 2\lambda\xi}{\pi^2} - \delta,\ \theta_2 + \frac{\nu + 2\lambda\xi}{\pi^2} + \delta\right]. \qquad (10)$$

5 Conclusions

We have derived limit bounds for the identification region for the vector of regression parameters in a linear regression model with interval-valued endogenous variable. We have discussed the case with two regression parameters, from where it is apparent which minimum assumptions are needed. However, we expect that the result could be generalized for a general number of regressors. The identification region for the vector of regression parameters is constructed as a limit of finite-sample identification regions for the OLS-estimator.

The derived bounds need not be tight. In [3, 4] it is shown that the finite-sample regions have a nice geometry; we conjecture that the geometry can be utilized in derivation of tighter bounds. Namely, we conjecture that the box (10) can be replaced by tighter convex set.

Finally, we tried to formulate as weak assumptions as possible on the process generating the interval-valued observations of the dependent variable. But, on the contrary, in many cases it might be plausible to admit *stronger* assumptions yielding *tighter* bounds compared to (9). Our model did not make any assumptions on the conditional distribution of y_i given the observed intervals $[\underline{y}_i, \overline{y}_i]$. In some applications, it might be plausible to assume that y_i is 'at least somehow, but reasonably random', even if one is not willing or able to make strong distributional assumptions. For example, it might be plausible to assume that $\mathrm{var}[y_i | \underline{y}_i, \overline{y}_i] \geq \alpha L_i^2$, where $L_i = \overline{y}_i - \underline{y}_i$ and $\alpha > 0$ is a universal constant. Such assumption (or other assumptions of this kind) might help in construction of tighter identification regions for the vector of regression parameters. Although such assumptions are not testable due to unobservability of y_i, they might be plausible in many applications since they restrict the class of possible conditional distributions $[y_i | \underline{y}_i, \overline{y}_i]$ only reasonably, ruling out some 'extreme' cases such as $\underline{y}_i = y_i$ for all i, or $\overline{y}_i = y_i$ for all i.

We studied the identification region for the vector of regression parameters as a limit of finite-sample identification regions for the OLS-estimator. Of course, analogous limit properties are of interest for other statistics, too. Namely, the limit behavior of the finite-sample regions for the variance of the disturbance term ε_i or usual goodness-of-fit measures would be desirable. Some finite-sample properties have already been studied in [5, 2, 8].

Acknowledgements

The work of M. Černý and M. Rada was supported by the Czech Science Foundation under Grant 16-00408S. The work of O. Sokol and V. Holý was supported by the Internal Grant Agency (I.G.A.) of University of Economics, Prague, Czech Republic, project F4/63/2016.

References

[1] Černý, M. and Hladík M.: The complexity of computation and approximation of the t-ratio over one-dimensional interval data. *Computational Statistics & Data Analysis* **80** (2014), 26–43.

[2] Černý, M. and Hladík, M.: Two optimization problems in linear regression with interval data, *Optimization* **66** (2017), 331–349.

[3] Černý, M., Hladík M. and Rada, M.: On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data, *Information Sciences* **244** (2013), 26–47.

[4] Černý, M. and Rada, M.: On the possibilistic approach to linear regression with rounded or interval-censored data, *Measurement Science Review* **11** (2011), 34–40.

[5] Černý, M. and Sokol, O.: Interval data and sample variance: A study of an efficiently computable case. In: Martinčík, D., Ircingová, J. and Janeček, P. (eds.): *Proceedings of 33rd Mathematical Methods in Economics* (2015), Cheb/Plzeň, University of West Bohemia, Czech Republic, 99–104.

[6] Manski, C.: *Partial Identification of Probability Distributions*. Springer Series in Statistics, Springer, 2003.

[7] Nguyen, H. T., Kreinovich, V., Wu, B. and Xiang, G.: *Computing Statistics under Interval and Fuzzy Uncertainty*. Edition: Studies in Computational Intelligence **393**, Springer, 2012.

[8] Sokol, O. and Rada, M.: Interval data and sample variance: How to prove polynomiality of computation of upper bound considering random intervals? In: Kocourek, A. and Vavroušek, M. (eds.): *Proceedings of 34th Mathematical Methods in Economics* (2016), Liberec, Czech Republic, 779–785.

[9] Stoye, J.: Partial identification of spread parameters, *Quantitative Economics* **1** (2010), 323–357.

[10] Tamer, E.: Partial identification in econometrics, *Annual Review of Economics* **2** (2010), 167–195.

Capacitated Vehicle Routing Problem Depending on Vehicle Load

Zuzana Čičková[1], Ivan Brezina[2], Juraj Pekár[3]

Abstract: The transport is one of the areas significantly contributing to the production of greenhouse emissions. An important part of the transport logistic is analyzing various routing problems and a variety of optimization models aimed at minimizing the total traveled distance are commonly known. When analyzing the fuel consumption, it is clearly observed that the traveled distance is not the only relevant factor contributing to the production of CO_2 emissions and undoubtedly also vehicle load has significant impact on its fuel consumption. The paper is aimed on a new vehicle routing problem that minimizes the fuel consumption, depending on the length of the traveled route and also on the vehicle load, namely capacitated vehicle routing problem depending on vehicle load (CVRPVL). The model will be presented in its nonlinear version, but also the mathematical formulation using only linear equations will be given (MIP formulation with linear equations). The difference in distribution compared to classical vehicle routing problem will be illustrated by solving specified problem in Slovakia.

Keywords: Capacitated Vehicle Routing Problem, Vehicle Load, CO_2 emissions
JEL Classification: C02, C61
AMS Classification: 90C11, 90B06

1 Introduction

A variety of optimization models to support decision making of distribution companies are commonly known. Distribution companies often implement models aimed at minimizing the total traveled distance. Widely used capacitated vehicle routing problem (CVRP), designs optimal set of routes of vehicles from origin (number of vehicles is not limited and each vehicle has the same capacity) aimed to serve a set of customers with a certain demand, where each vehicle travels exactly one route so that the demand of customers must to be met in full by exactly one vehicle and capacity must not be exceeded. It is assuming the known lowest cost (usually distance) between origin and each customer, as well as between each pairs of customers ([1], [3], [4], [5], [7], [8]). The CVRP is an important combinatorial optimisation problem, e.g. [9] describes several case studies where the application of CVRP has led to substantial cost savings.

However, when analyzing the fuel consumption, it is clearly observed that the traveled distance is not the only relevant factor and undoubtedly also vehicle load has significant impact on the its consumption. The paper is aimed on a model that minimizes the fuel consumption, depending on the length of the traveled route and also on the vehicle load, which we named **capacitated vehicle routing problem depending on vehicle load** (CVRPVL). The paper is structured as follows: In the first part of second section we present presuppositions and a mathematical model of CVRPVL. Because of non-linearity in objectives and also in the structural equation, we propose the modified formulation of the model in the second part of that section. Although it is binary programming problem, it contains only linear equations and linear objective. The third part is devoted to an illustrative example containing cities in Slovakia, where we calculate optimum route based on CVRP as well as on a CVRPVL. The results illustrate the difference using both of approaches, while we also report an achieved decrease in the fuel consumption.

2 Capacitated Vehicle Routing Problem Depending on Vehicle Load

Let us introduce the mathematical formulation of capacitated vehicle routing problem depending on vehicle load (CVRPVL) based on famous Miller-Tucker-Zemlin's formulation of the traveling salesman problem ([6]).

[1] Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1, 852 35 Bratislava, e-mail: zuzana.cickova@euba.sk
[2] Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1, 852 35 Bratislava, e-mail: ivan.brezina@euba.sk.
[3] Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemská cesta 1, 852 35 Bratislava, e-mail:, juraj.pekar@euba.sk.

Following notation is used: Let $N = \{1, 2, ...n\}$ be the set of served nodes (customers) and let $N_0 = N \cup \{0\}$ be a set of nodes that represents the customers as well as the origin (depot). Further on there exists a matrix $D_{(n+1)x(n+1)}$ associated with pairs $i, j \in N_0$, $i \neq j$ that represents the minimum distances between all the pairs of nodes (customers and the depot). In general, one way how to mathematically describe routing problems is using binary variables x_{ij} $(i, j \in N_0, i \neq j)$ that enable to model if the node i precedes node j in a route of the vehicle $x_{ij} = 1$ and $x_{ij} = 0$ otherwise. Certain demand q_i, $i \in N$ is associated with each customer. All the demands have to be met from the initial node ($i = 0$) in such a way that the distribution is performed using a vehicles with a certain capacity (g) (the number of vehicles is not limited). The model implicitly assumes that $q_i \leq g$ for all $i \in N$, i.e. the demand of each customer does not exceed the capacity of the vehicle. Further on, the variables $u_i, i \in N$ are used. Those variables represent cumulative demand of customers on one particular route. Now consider new known parameters associated with the fuel consumption. Let parameter c_0 be the vehicle consumption per unit distance and let c_1 be parameter representing the increase in consumption per unit distance for one unit of vehicle load. The goal of CVRPVL is to establish such distribution, which minimizes whole vehicle's fuel consumption (not minimizing the distance as in classical CVRP), but also consider the CVRP's restrictions: the origin represents initial node and also the final node of every route, from this node the demands q_i, $i \in N$ of all the other nodes are met (in full), each node (except origin) is visited exactly once and total demand on route must not exceed the capacity of the vehicle (g).

Let us recapitulate the model parameters as well as model variables more clearly:

Parameters: n – number of customers (served nodes),

$N = \{1, 2, ...n\}$ – set of customers (served nodes),

$N_0 = N \cup \{0\}$ – set of customers and the origin,

d_{ij}, $i, j \in N_0$, $i \neq j$ – shortest distance moving from node i to node j,
q_i, $i \in N$ – demand of i-th customer,
g – capacity of vehicles,
c_0 – vehicle consumption per unit distance
c_1 – increase in consumption per unit distance for one unit of vehicle load.

Variables: x_{ij}, $i, j \in N_0$, $i \neq j$ representing if the node i precedes node j in a final route ($x_{ij} = 1$) or not ($x_{ij} = 0$), it should be noted that the resulting routes depend on the type of the problem. When goods are collected than the optimal route represented by variables $x_{ij} = 1$, is the route from i-th node to j-th node. In the case of distribution those variables represent the route from j-th to i-th node.

$u_i, i \in N$ based on Miller-Tucker- Zemlin's formulation, but representing vehicle load; in the case of goods collection it is current load of vehicle on its route from the i-th node, but if the goods are distributed it is vehicle's load on its route to the i-th node.

The model describing CVRPVL deals with these variables:

$$x_{ij} \in \{0,1\},\ i, j \in N_0, i \neq j \tag{1}$$

$$u_i \geq 0,\ i \in N,\ u_0 = 0 \tag{2}$$

where the objective can be written as follows:

$$\min f(\mathbf{X}, \mathbf{u}) = \sum_{i \in N_0} \sum_{\substack{j \in N_0 \\ i \neq j}} (c_0 + c_1 \cdot u_i) d_{ij} x_{ij} \tag{3}$$

subject to

$$\sum_{i \in N_0} x_{ij} = 1,\ j \in N,\ i \neq j \tag{4}$$

$$\sum_{j \in N_0} x_{ij} = 1,\ i \in N,\ i \neq j \tag{5}$$

$$x_{ij}(u_i + q_j - u_j) = 0,\ i \in N_0,\ j \in N,\ i \neq j \tag{6}$$

$$u_i \leq g,\ i \in N \tag{7}$$

The objective function determines whole vehicle's fuel consumption: minimizing the expression $\sum_{i\in N_0}\sum_{\substack{j\in N_0\\ i\neq j}} c_0 d_{ij} x_{ij}$ enables modeling the fuel consumption depending on the traveled route and expression $\sum_{i\in N_0}\sum_{\substack{j\in N_0\\ i\neq j}} c_1 u_i d_{ij} x_{ij}$ enables modeling increased consumption depending on the vehicle's load. Binary variables (1) give equations (4) to (7) standard meanings: equations (4) and (5) ensure that each customer (except the origin) is visited exactly once. Equations (6) are anti-cyclical conditions that prevent the formation of such sub-cycles which do not contain an origin and they also enable calculating values of variables $u_i,\ i \in N$ (see also equations (2) and (7)).

The model of CVRPVL (3)-(7) contains also nonlinear expressions; the nonlinear objective function (3) and nonlinear equation (6) that may complicate the possibility of solution of related problems (MINLP type). The better way to solve such problems is to use a mixed integer programming (MIP) formulation (with linear objective function and constraints), which extends the possibilities of the solution running standard software for solving MIP problems. Let's us provide the linearization. Except the variables (1) and (2) we will use also the non-negative variables

$$n_{ij} \geq 0,\ i,\ j \in N_0, i \neq j \tag{8}$$

with the following notation: in the case of goods collection they represent load of vehicle on its route from *i*-th to *j*-th node, but if the goods are distributed it is load of vehicle on its route from *j*-th to *i*-th node (please note that the meaning of binary variables x_{ij} also differ depending on problem type (collection or distribution)).

Now let us make the objective function (3) linear:

$$\min f(\mathbf{X},\mathbf{u},\mathbf{n}) = c_0 \sum_{i\in N_0}\sum_{\substack{j\in N_0\\ i\neq j}} d_{ij} x_{ij} + c_1 \sum_{i\in N_0}\sum_{\substack{j\in N_0\\ i\neq j}} d_{ij} n_{ij} \tag{9}$$

Thus the first part of the sum represents the fuel consumption dependence on the traveled route and its second part enables modeling increased consumption dependence on the vehicle's load. To allow calculation of variables n_{ij} we need to introduce new equations:

$$u_i \leq n_{ij} + (1 - x_{ij})\, g,\ \ i,\ j \in N_0,\ i \neq j \tag{10}$$

At last make the equation (6) linear. The linearization is very simple:

$$u_i + q_j \leq u_j + (1 - x_{ij})g,\ i \in N_0, j \in N,\ i \neq j \tag{11}$$

In both cases the parameter g (capacity of the vehicles) is used as a tool that enables the calculation of variables u_i and n_{ij} if the variable $x_{ij} = 1$ and it also ensures feasibility of solution in the case that $x_{ij} = 0$.

3 Illustrative Example

Consider scheduling in network consisting of origin (0 – Žilina) from where 8 nodes (customers) need to be served (1 – Košice, 2 – Banská Bystrica, 3 – Bratislava, 4 – Nitra, 5 – Prešov, 6 – Trenčín, 7 – Trnava). Values of input parameters were set as follows:

$n = 8,\ N = \{1, 2, ...7\},\ N_0 = N \cup \{0\}$ - number and sets of nodes (customers and also origin),

$\mathbf{q} = (8, 4, 10, 10, 9, 8, 9)^{\mathrm{T}}$ - vector of customers' demands,

$g = 24$ – capacity of vehicles,

$c_0 = 0.22$ – vehicle consumption per unit distance (liter/km),

$c_1 = 0.007$ – increased consumption per unit distance and per unit of load (liter/tonne/km),

$\mathbf{D} = \{d_{ij}\},\ i, j \in N_0,\ i \neq j$ – matrix of shortest distances from node *i* to node *j*

$$\mathbf{D} = \begin{pmatrix} 0 & 230 & 89 & 200 & 140 & 221 & 73 & 151 \\ 230 & 0 & 213 & 391 & 302 & 35 & 303 & 349 \\ 89 & 213 & 0 & 208 & 119 & 195 & 142 & 166 \\ 200 & 391 & 208 & 0 & 88 & 403 & 127 & 47 \\ 140 & 302 & 119 & 88 & 0 & 315 & 85 & 46 \\ 221 & 35 & 195 & 403 & 315 & 0 & 293 & 361 \\ 73 & 303 & 142 & 127 & 85 & 293 & 0 & 78 \\ 151 & 349 & 166 & 47 & 46 & 361 & 78 & 0 \end{pmatrix}$$

Firstly we solve the problem using linearized model of CVRPVL. For the comparison the solution using model of CVRP ([2]) is provided. Both models are implemented in software GAMS on PC with Intel ® Core ™ i7-3770 CPU with a frequency of 3.40 GHz and 8 GB of RAM under MS Windows 8. Results are given in Table 1 (for CVRPVL) and in Table 2 (for CVRP).

Route	**Sequence**	**Distance**	**Fuel consumption depending on load**
Route	0-7-3-0-6-4-2-0-5-1-0	1250	349.536
Route 1	0-7-3-0	398	110.933
Route 2	0-6-4-2-0	366	103.424
Route 3	0-5-1-0	486	135.179

Table 1 CVRPVL. *Source:* Own compilation.

Route	**Sequence**	**Distance**	**Fuel consumption depending on load**
Route	0-7-3-0-6-4-0-1-5-2-0	1245	354.631
Route 1	0-7-3-0	398	110.933
Route 2	0-6-4-0	298	80.708
Route 3	0-1-5-2-0	549	162.99

Table 2 CVRP. *Source:* Own compilation.

The change in resulting route length is observed, when we provide two different models (CVRPVL and CVRP). In the case of implementing CVRP we obtain the value of the total traveled distance 1245 *km*. Using model CVRPVL leads to the route length of 1250 *km*. But when we calculate the fuel consumption with dependence on vehicle load, resulting values are those: 354.631 *l* in the case of CVRP and 349.536 *l* in the case of CVRPVL. Using CVRPVL model decreased fuel consumption by 1.44 percent. Change can be clearly attributed to rearranging of nodes to fulfill the goal of minimizing the fuel consumption with respect to the vehicle load which results in different routes structure.

4 Conclusion

This paper considers the modification of CVRP, which we named capacitated vehicle routing problem depending on vehicle load (CVRPVL). Modifications of the classical CVRP are usually aimed to find a minimal cost of routes. Authors constructed a model that enables to minimize CO_2 production taking into account not only traveled distance but also the weight of the loaded goods. The first part of the paper is devoted to presupposition and non-linear mathematical model of CVRPVL and also to its modified version containing only linear objective as well as linear constraints. The next section gives an example of using CVRPVL model, while justifying the difference compared to classical CVRP. Also the results of given example show that when solving practical problems it is necessary to include the weight of the load, since that parameter largely affects the individual

routes. Obviously a CO_2 production in a real life distribution is affected by more factors: route gradient, weather situation, etc. Modeling such generalized situation is open for the future research.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic – VEGA, grant no. 1/0245/15 „Transportation planning focused on greenhouse gases emission reduction“.

References

[1] Čičková, Z., Brezina, I., Pekár, J.: Solving the routing problems with time windows. In: *Self-organizing migrating algorithm : methodology and implementation* (editors: Donald Davendra, Ivan Zelinka.). Springer International Publishing AG Switzerland, Cham, 2016, 207-236.

[2] Čičková, Z., Brezina, I., Pekár, J.: Routes design using own and hired vehicles. In: *Mathematical methods in economics 2015*. University of West Bohemia, Cheb, 2015, 105-108.

[3] Eksioglu, B., Vural, A. V., Reisman, A.: The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering* **57**(4/2009), 1472–1483.

[4] Fábry, J., Kořenář, V., Kobzareva, M.: Column Generation Method for the Vehicle Routing Problem – The Case Study. In: *Mathematical Methods in Economics 2011*. PROFESSIONAL PUBLISHING, Praha, 2011, 140–144.

[5] Golden, B. L., Raghavan, S., Wasil, E. A.: *The Vehicle Routing Problem: Latest Advances and New Challenges.* Springer, New York, 2008.

[6] Miller, C.E., Tucker, A.W., Zemlin, R.A.: Integer programming Formulation of Traveling Salesman Problems. *Journal of the ACM* **7**(4/1960), 326-329.

[7] Palúch, S., Peško, Š.: *Kvantitatívne metódy v logistike*. EDIS vydavateľstvo ŽU, Žilina, 2006.

[8] Pekár, J., Brezina, I., Čičková, Z.: Synchronization of capacitated vehicle routing problem among periods. *Ekonomický časopis* **65** (1/2017), 66-78.

[9] Toth, P., Vigo, D.: *The Vehicle Routing Problem (Monographs on Discrete Mathematics and Applications)*. SIAM. Philadelphia, 2002.

Modeling Unemployment Rate in Spain: Search and Matching Approach

Ondřej Čížek[1]

Abstract. The goal of the paper is to model unemployment rate in Spain by using a methodology of search and matching models. Econometric estimation based on Bayesian techniques is performed and empirical performance of the model is evaluated. The estimated model is able to replicate key features of the data with the exception of autocorrelation functions of the observed variables. Matching autocorrelation coefficients of higher order is especially problematic and the model underestimates these coefficients. The unemployment rate in Spain is very persistent and can be far away from a steady state for a very long time. Underestimation of higher-order autocorrelation coefficients suggests that the basic search and matching model has difficulties in capturing this aspect of unemployment rate in Spain. Forecasting performance of the model was also analyzed by computing ex post dynamic one-year-ahead forecasts of unemployment rate. The important result is that forecasting performance of the model during the crisis periods was even better than for the periods before the current economic crisis.

Keywords: Search-matching model, unemployment rate, market tightness, Bayesian estimation.

JEL Classification: C51, E24, J60
AMS Classification: 91G70

1 Introduction

The unemployment rate in Spain during the current economic crisis has become a serious economic problem. Some of the recent empirical studies analyzing the Spanish labor market are as follows. Gender differences in the Spanish unemployment is studied by De la Rica, Rebollo-Sanz [4]. The relation between unemployment, wages and productivity in Spain is studied by Pisa, Sánchez [13]. Bentolila et al. [3] investigate unemployment in Spain during the current economic crisis and make a comparison with France. De Toledo et al. [5] perform an empirical analysis of the matching process in the Spanish labor market. Garcia-Pepez [7] models the Spanish labor market using the search and matching modeling framework.[2] Schwarzmüller, Stähler [15] apply search and matching methodology within DSGE (Dynamic Stochastic General Equilibrium) framework to analyze labor market reforms in Spain. Altuzarra [1] applies a battery of unit root tests and finds that unemployment in Spain is highly persistent.

The aim of this paper is to model unemployment by applying standard search and matching theory of labor market which dominates the economic literature. Specifically, the model developed by Hagedorn, Manovskii [8] (HM model, hereafter) is applied for this purpose as this model has attained a great attention in the economic literature. Mostly, search and matching models are only calibrated. This paper contributes to the literature by econometric estimation of this model which is based on advanced Bayesian techniques. The significant finding is that the HM model has difficulties in capturing highly persistent behavior of unemployment rate in Spain.

The structure of the paper is as follows. The chapter 2 briefly summarizes the model equations. Data are described in the chapter 3 and econometric estimation is discussed in the subsequent chapter 4. Empirical performance of the model is evaluated in the chapter 5 and the final chapter 6 presents a conclusion.

2 Model

For convenience, the key equations of the HM model are briefly summarized in this chapter. Output per worker is denoted by p_t and follows the first-order autoregressive process

$$\log(p_t) = \rho^p \cdot \log(p_{t-1}) + \varepsilon_{t,1}, \tag{1}$$

[1] University of Economics in Prague, Department of Econometrics, W. Churchill Sq. 4, 130 67, Prague 3, Czech Republic, cizeko@vse.cz.

[2] The search and matching modeling framework is summarized in a textbook treatment by the Nobel Price winner Pissarides [14].

where $\rho^p \in (0,1)$ and $\varepsilon_{t,1} \sim N(0, \sigma_1^2)$ is i.i.d. productivity shock.

Flow cost c_t of posting a vacancy is assumed to change over the business cycle according to

$$c_t = c_K \cdot p_t + c_W \cdot p_t^{\xi}. \tag{2}$$

Let u_t denote number of unemployed, $n_t = 1 - u_t$ number of employed[3], v_t number of vacancies and $\theta_t = v_t / u_t$ market tightness. The matching process is modeled by standard Cobb-Douglas matching function[4]

$$m(u_t, v_t) = m_0 \cdot u_t^{\eta} \cdot v_t^{1-\eta} \cdot \exp(\varepsilon_{t,2}), \tag{3}$$

where the shock to matching efficiency $\varepsilon_{t,2}$ is supposed to be persistent

$$\varepsilon_{t,2} = \rho^m \cdot \varepsilon_{t-1,2} + \tilde{\varepsilon}_{t,2}, \tag{4}$$

where $\rho^m \in (0,1)$ and $\tilde{\varepsilon}_{t,2} \sim N(0, \sigma_2^2)$ is i.i.d. shock.

The probability for an unemployed worker to be matched with a vacancy equals $f(\theta_t) = m(u_t, v_t)/u_t$ and the probability for a vacancy to be filled is $q(\theta_t) = m(u_t, v_t)/v_t$. Workers and firms separate with a constant[5] probability s per period. Evolution of employment rate is given by

$$n_{t+1} = (1-s) \cdot n_t + f_t \cdot u_t + \varepsilon_{t+1,3}, \tag{5}$$

where $\varepsilon_{t+1,3} \sim N(0, \sigma_3^2)$ is i.i.d. shock to the process of unemployment.

It can be shown by standard methods that the first-order conditions lead to the equilibrium condition of the form

$$c_t / (\delta \cdot q(\theta_t)) = E_t\left[(1-\beta) \cdot (p_{t+1} - z) - c_{t+1} \cdot \beta \cdot \theta_{t+1} + (1-s) \cdot c_{t+1} / q(\theta_{t+1})\right]. \tag{6}$$

Wages are determined by the generalized Nash bargaining solution

$$w_t = \beta \cdot p_t + (1-\beta) \cdot z + c_t \cdot \beta \cdot \theta_t + \varepsilon_{t,4}, \tag{7}$$

where $\beta \in (0,1)$ is the bargaining power of workers and $\varepsilon_{t,4} \sim N(0, \sigma_4^2)$ is i.i.d. shock added for the purpose of econometric estimation. Unemployed workers get a flow utility z from leisure/non-market activity.

3 Data

The source of the data is OECD database. All data were seasonally adjusted. The first observable variable is the standardized unemployment rate in Spain u_t from 1986 M4 to 2016 M8. The second variable the market tightness θ_t from 1986 M4 to 2005 M4 calculated as a ratio of number of unfilled vacancies to number of unemployed persons. The third observable variable is productivity p_t which is measured as a relative deviation from a linear trend of the industrial production index in manufacturing from 1986 M4 to 2016 M8. The last observable variable relates to wages w_t. The variable w_t is measured as a relative deviation from a linear trend of the index of (real)

[3] Labor force is normalized to one which is common practice within search and matching modeling framework. Number of (un)employed is thus also interpreted as an (un)employment rate.
[4] Hagedorn, Manovskii [8] applied another form of matching function. Nonetheless, standard Cobb-Douglas matching function performed better from an empirical point of view.
[5] See Hall [9] or Shimer [16] for the empirical evidence that fluctuations in job finding probability during business cycle frequencies are substantial, while separation probability is nearly acyclic.

hourly earnings in manufacturing. This measure of w_t implies that its mean value equals approximately to one as in the case of p_t. Therefore, the monthly rate of change $(w_t - w_{t-1})/w_{t-1}$ is used as the observable variable. The index of hourly earnings in manufacturing was transformed from quarterly frequency into a monthly frequency by cubic spline. The data ranges from 1986 M4 to 2016 M4 after this transformation.

4 Econometric estimation

The econometric estimation is based on Bayesian econometric techniques and is performed using the Matlab toolbox Dynare (version 4.4.3). The priors for the baseline model are reported in the following table 1.

Parameter	Description	Density	Mean	Std. Dev.
δ	discount factor	fixed	$0.99^{1/3}$	-
s	separation probability	fixed	0.014	-
ρ^p	AR coef. in productivity process	beta	0.80	0.20
ρ^m	AR coef. in matching process	beta	0.80	0.20
c_K	vacancy cost	beta	0.47	0.20
c_w	vacancy cost	beta	0.11	0.20
ξ	vacancy cost	beta	0.45	0.20
β	workers' bargaining power	beta	0.50	0.20
η	elasticity of matching	beta	0.50	0.20
z	value of non-market activity	beta	0.40	0.20
$\sigma_{1,2,3,4}$	std. dev. of shocks	inv.gamma	0.01	1

Table 1 Parameter description and prior densities

The results of the estimation are presented in the form of posterior means together with 90% confidence intervals:

Parameter	Posterior mean	90% confidence interval
ρ^p	0.830	(0.822, 0.839)
ρ^m	0.958	(0.940, 0.977)
c_K	0.974	(0.951, 0.996)
c_w	0.981	(0.961, 1.000)
ξ	0.842	(0.705, 0.978)
β	0.014	(0.006, 0.021)
η	0.973	(0.948, 0.997)
z	0.9853	(0.9846, 0.9861)
σ_1	0.025	(0.024, 0.027)
σ_2	0.117	(0.108, 0.127)
σ_3	0.0030	(0.0028, 0.0032)
σ_4	0.015	(0.014, 0.016)

Table 2 Econometric estimates of the coefficients

The interesting result relates to the parameters z and β. The estimated posterior mean of these parameters is 0.9853 and 0.014 which is close to the calibrated values $z = 0.955$, $\beta = 0.052$ used in the paper by Hagedorn, Manovskii [8]. These authors showed that their calibration strategy of z close to 1 and β close to zero generates

volatilities of unemployment and market tightness similar to that observed in the U.S. data. The results of the econometric estimation thus suggest that their calibration strategy might be appropriate not only for the U.S. data.

5 Empirical performance of the model

5.1 Matching moments

Firstly, selected moments of market tightness and unemployment rate will be compared in order to assess empirical performance of the model.

		market tightness	**unemployment rate**
mean	observed data	0.037	0.169
	model	0.037	0.150
standard deviation	observed data	0.027	0.051
	model	0.018	0.044

Table 3 Comparison of moments calculated from observed data and from data generated by the model

The correlation of market tightness and unemployment rate in the observed data was (-0.715) which is exactly the same as the correlation of these variables in the data generated by the model. From these facts, it is evident that the model is able to reproduce selected features of the observed data. Nonetheless, the model fails to match autocorrelation functions which is documented in the following figure.

Figure 1 Comparison of autocorrelation functions

Matching autocorrelation coefficients of higher order is especially problematic and the model underestimates these coefficients. This finding suggests that the selected variables are much more persistent. Unemployment rate in Spain can be far away from a steady state for a very long time and the HM model seems to have problems to generate data with this property.

5.2 Forecasting performance

Forecasting performance of the model is evaluated by calculating ex post dynamic one-year-ahead forecasts of the unemployment rate. The information about u_t, p_t and $\varepsilon_{t,2}$ is used together with the knowledge of the policy function coefficients to make a dynamic forecast 12 months ahead $u_{t+12|t}$ which is then compared to the observed value u_{t+12}. This exercise is repeated for $t = 1986\ \text{M4},...,2015\ \text{M8}$. Smoothed value of the variable $\varepsilon_{t,2}$ was utilized as it is not directly observable. Comparison of forecasts $u_{t+12|t}$ with observed values u_{t+12} is depicted at the figure.

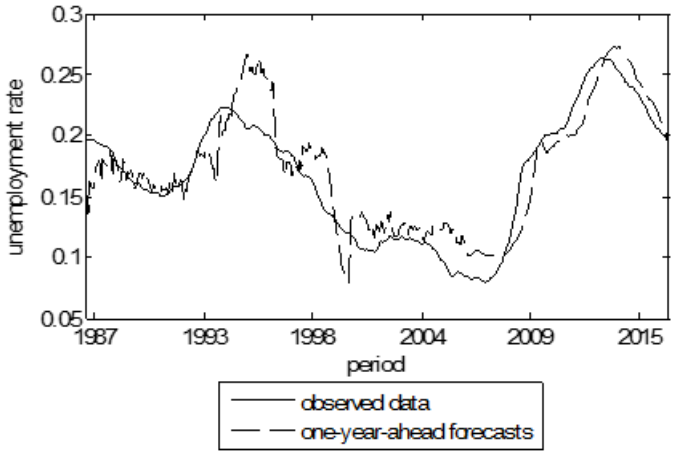


Figure 2 Dynamic ex post one-year-ahead forecasts of the unemployment rate

The figure documents that forecasting performance of the model is satisfactory. The mean squared error (MSE) of the forecasts is 5.1738e-04. The interesting result is that forecasting performance for the periods of the economic crisis (from 2008) is even better than for the periods before the crisis.

6 Conclusion

The unemployment rate in Spain was modeled by applying search and matching approach. The famous HM model was applied for this purpose. The important result is that econometric estimation of this model supported the calibration strategy suggested by Hagedorn, Manovskii [8]. Specifically, the value of non-market activity was estimated close to one and the estimate of bargaining power of workers turned out to be close to zero. The calibration strategy proposed by Hagedorn and Manovskii thus seems to be appropriate not only for the U.S. data.

The empirical performance of the model was also analyzed. The analysis showed that the estimated model is able to generate data that have mean values and standard deviations very close to that observed in the dataset. Nonetheless, the model is not able to match autocorrelation functions of the most important variables like unemployment rate and market tightness. It was found that the model systematically underestimates (higher order) autocorrelation coefficients. The persistence observed in the real data is very high. For this reason, unemployment rate might be far away from a steady state for quite a long time. The estimated HM model seems to have difficulties to generate data with this property. Forecasting performance was also studied and found to be satisfactory especially for the periods of the current economic crisis.

The empirical analysis of the labor market performed in this paper could be extended in several dimensions. Two promising directions of research are as follows. Firstly, unemployment rate could be disaggregated and studied separately for different groups of workers (Janíčko [10]). Secondly, standard search and matching model of the labor market could be combined with a DSGE approach (Němec [12], Galí, Smets, Wouters [6], Krause, Lubik [11], Trigari [17], Bouda, Formánek [2]).

Acknowledgements

Financial support of VŠE IGA IG403036 is gratefully acknowledged by the author. Paper was processed with contribution of long term support of scientific work on Faculty of Informatics and Statistics, University of Economics, Prague (IP 400040).

References

[1] Altuzarra, A.: Measuring Unemployment Persistence by Age and Gender. *Romanian Journal of Economic Forecasting* **18**(4) (2015), 110-133.

[2] Bouda, M., and Formánek, T.: Housing Sector-specific DSGE Model with Applications to Czech and Slovak Economies. *Ekonomický časopis* **64**(8) 2014, 805–822.

[3] Bentolila, S., Cahuc, P., Dolado, J. J., and Le Barbanchon, T.: Two-Tier Labour Markets in the Great Recession: France versus Spain. *Economic Journal* **122**(562) (2012), F155-F187.

[4] De la Rica, S., and Rebollo-Sanz, Y. F.: Gender Differentials in Unemployment Ins and Outs during the Great Recession in Spain. *Economist-Netherlands* **165**(1) 2017, 67-99.
[5] De Toledo, P. A., Nunez, F., and Usabiaga, C.: An Empirical Analysis of the Matching Process in Andalusian Public Employment Agencies. *Hacienda Publica Espanola* **198** (2011), 67-102.
[6] Galí, J., Smets, F. R., and Wouters, R.: Unemployment in an Estimated New Keynesian Model. In: NBER Macroeconomics Annual 2011 (Acemoglu, D., and Woodford, M., eds.). University of Chicago Press, 2012, 329-360.
[7] Garcia-Pepez, J. I.: Job Separation in a Non-Stationary Search Model: A Structural Estimation to Evaluate Alternative Unemployment Insurance Systems. *Journal of Applied Econometrics* **21**(2) (2006), 245-272.
[8] Hagedorn, M., and Manovskii, I.: The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited. *American Economic Review* **98**(4) (2008), 1692-1706.
[9] Hall, R. E: Job Loss, Job Finding and Unemployment in the U.S. Economy over the Past Fifty Years. In NBER Macroeconomics Annual 2005 (Gertler, M., and Rogoff, K., eds.). Cambridge: MIT Press, 2005, 101-131.
[10] Janíčko, P.: The Vulnerable Groups on the Czech Labour Market: Trends, Actual State and Influence of Economic Policies. In: 6th International Days of Statistics and Economics (Löster, T., and Pavelka, T., eds.), Prague, Czech Republic, 2012, 470-480.
[11] Krause, M. U., and Lubik, T. A.: The (Ir)Relevance of Real Wage Rigidity in the New Keynesian Model with Search Frictions. *Journal of Monetary Economics* **54**(3) (2007), 706-727.
[12] Němec, D.: Investigating Differences between the Czech and Slovak Labour Market Using a Small DSGE Model with Search and Matching Frictions. *Czech Economic Review* **7** (2013), 21-41.
[13] Pisa, M., and Sánchez, R.: Work Incentive and Productivity in Spain. *Prague Economic Papers* **25**(1) (2016), 99-111.
[14] Pissarides, C. A.: *Equilibrium Unemployment Theory*. 2nd ed. MIT Press, 2000.
[15] Schwarzmüller, T., and Stähler, N.: Reforming the Labor Market and Improving Competitiveness: An Analysis for Spain Using FiMod. *Series-Journal of the Spanish Economic Association* **4**(4) (2013), 437-471.
[16] Shimer, R.: Reassessing the Ins and Outs of Unemployment. *Review of Economic Dynamics* **15**(2) (2012), 127-148.
[17] Trigari, A.: The Role of Search Frictions and Bargaining for Inflation Dynamics. IGIER Working Paper no. 304, Innocenzo Gasparini Institute for Economic Research, Milano, 2006.

Correlation dimension as a measure of stock market variability

Martin Dlask[1], Jaromir Kukal[2]

Abstract. Economical time series often show fractional behaviour. Since this type of data can be considered as a realisation of stochastic process with unknown properties, it can be analysed with the tools of fractal geometry. In the particular case of stock market indices, one can investigate any individual index using state space reconstruction according to Whitney theorem. The second possibility of the analysis is the study of development of a group of different indices describing the state of the market at a specific time. Subsequently, long-time stock market history is useful for the investigation of the states as vectors in Euclidean space. Their fractal nature can be studied using correlation dimension. The paper presents several approaches to its estimation and compares the results in terms of the individual stock market behaviour. As a referential technique, the classical approach using correlation sum is presented and its performance is discussed in the context of obtained results. Furthermore, the analysis is useful for the dependency analysis and measure of predictability of time series.

Keywords: time series, state space reconstruction, correlation dimension, stock market indices

JEL classification: E44
AMS classification: 60G22, 62M10

1 Introduction

The correlation dimension D_2 is a particular case of Renyi dimension [9, 10] D_α for $\alpha \geq 0$ that is defined based on distances between points in the investigated set in $\mathbf{R}^n$. Together with the capacity and information dimension, it belongs to the group of entropy-based dimensions. Methods that are used to estimate the Renyi dimension are usually different for different parameter α.

Stock market indices that will be later investigated in this chapter can be considered as a realisation of a random process. There are generally two tools of fractal geometry how to analyse time series. The first approach estimates its Renyi or Hausdorff dimension D based on the trajectory of the index. The second approach employs fractional processes [17] and estimates the Hurst exponent H. When the set fulfils the open set condition, the Hausdorff dimension equals Renyi dimension and the Hurst exponent equals $H = 2 - D$.

The traditional way how to estimate correlation dimension was presented by Grassberger and Proccacia [12, 13] and is used till today in the applications where the extensive precision is not necessary. For instance, the original approach can be used in biomedicine for electroencephalography signal analysis [19] or in cardiology [15].

Recently there were attempts to enhance the accuracy of correlation dimension estimation based on the first derivative of Gaussian kernel correlation sum [5] or utilizing the self similarity property and clustering [21]. These methods can be subsequently used for example for the analysis of finance market under the fractional Brownian motion assumption [18].

Stock market indices are popular object for the fractal investigation. There are several papers dealing with their analysis using traditional singular value decomposition entropy [14], Hurst exponent [3, 6, 8] or geometrical properties of fractal sets [1]. The main benefit of precise dimension estimation is the information of the dependency of the time series that can be later used for predictions.

In this paper we describe the traditional correlation dimension that leads to biased estimate of D_2 and we suggest a spectral method that can provide more exact estimate of the correlation dimension.

[1]Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering, Trojanova 12, Praha, martin.dlask@fjfi.cvut.cz
[2]Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering, Trojanova 12, Praha, jaromir.kukal@fjfi.cvut.cz

2 Correlation Dimension

Correlation dimension, introduced by Grassberger and Procaccia [12, 13], involves measuring the distance between all pairs of points in the investigated set. For the Lebesgue measurable set $\mathcal{F} \subset \mathbb{R}^n$, the *correlation sum* [12] is defined for any distance between points $r > 0$ as the limit case

$$\mathrm{C}(r) = \lim_{N\to\infty} \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathrm{I}(\|\boldsymbol{x_i} - \boldsymbol{x_j}\| \leq r), \tag{1}$$

where $\|.\|$ denotes a Euclidean norm that is rotation invariant, I is the indicator function and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are vectors from $\mathcal{F}$. Because the correlation dimension expresses the relative amount of points whose distance is less than r, the correlation sum can be rewritten as

$$\mathrm{C}(r) = \mathop{\mathrm{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\mathrm{U}(\mathcal{F})} \mathrm{I}(\|\mathbf{x} - \mathbf{y}\| \leq r) = \mathop{\mathrm{prob}}_{\boldsymbol{x},\boldsymbol{y}\sim\mathrm{U}(\mathcal{F})} (\|\boldsymbol{x} - \boldsymbol{y}\| \leq r)\,, \tag{2}$$

for $\boldsymbol{x}, \boldsymbol{y}$ that are uniformly distributed on $\mathcal{F}$. Therefore, $\mathrm{C}(r)$ is a cumulative distribution function of random variable $r = \|\boldsymbol{x} - \boldsymbol{y}\|$. The correlation dimension D_2 of set $\mathcal{F}$ is based on the correlation sum and is defined as

$$D_2 = \lim_{r\to 0^+} \frac{\ln \mathrm{C}(r)}{\ln r}, \tag{3}$$

if the limit exists.

The definition (3) cannot be used directly for the estimation of correlation dimension from finite sample size. Therefore, one usually employs linear regression model

$$\ln C(r) = A + D_2 \cdot \ln r \tag{4}$$

for small values of r. The standard deviation of the estimate gets smaller with the increasing number of data points, however, the estimate is usually biased by its nature [16].

3 Rotational Spectrum

The goal of the method is to obtain a one-dimensional function as a derivative of the power spectrum, which is useful in fractal analysis. This method is described in [7] in detail. The procedure was inspired by Debye [4] and by his X-ray diffraction method, which is often referred to as the Debye-Scherrer method. We denote $\mathrm{SO}(n)$ as the group of all rotations in $\mathbb{R}^n$ around the origin. Because any rotation $\mathrm{R} \in \mathrm{SO}(n)$ is a linear transform, the following equation holds

$$\mathrm{R}(\boldsymbol{x}) - \mathrm{R}(\boldsymbol{y}) = \mathrm{R}(\boldsymbol{x} - \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| \cdot \boldsymbol{\xi}, \tag{5}$$

where $\boldsymbol{\xi}$ is a direction vector satisfying $\|\boldsymbol{\xi}\| = 1$ and $\boldsymbol{\xi} \in \mathcal{S}_{n-1}$ for an n-dimensional sphere

$$\mathcal{S}_{n-1} = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| = 1\}\,. \tag{6}$$

Using the factorisation of angular frequency $\boldsymbol{\omega} = \Omega \cdot \boldsymbol{\psi}$ for $\Omega \in \mathbb{R}_0^+$ and normalisation vector $\boldsymbol{\psi} \in \mathcal{S}_{n-1}$, we can define *rotational spectrum* as

$$\mathrm{S}(\Omega) = \mathop{\mathrm{E}}_{\mathrm{R}\in\mathrm{SO}(n)} \mathop{\mathrm{E}}_{\boldsymbol{\psi}\in\mathcal{S}_{n-1}} \mathop{\mathrm{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\mathrm{U}(\mathcal{F})} \exp(-\mathrm{i}\Omega\boldsymbol{\psi}\mathrm{R}(\boldsymbol{x} - \boldsymbol{y})). \tag{7}$$

The rotational spectrum can be expressed analytically as

$$\mathrm{S}(\Omega) = \mathop{\mathrm{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\mathrm{U}(\mathcal{F})} \mathrm{H}_n(\Omega\,\|\boldsymbol{x} - \boldsymbol{y}\|), \tag{8}$$

for function $\mathrm{H}_n(q)$ dependent on the dimension of the rotation as

$$\mathrm{H}_n(q) = \frac{2^{\frac{n-2}{2}} \cdot \Gamma\left(\frac{n}{2}\right)}{q^{\frac{n-2}{2}}} \mathrm{J}_{\frac{n-2}{2}}(q). \tag{9}$$

Taking infinite-dimensional rotation, the the kernel function H_n can be expressed in the form of Gaussian function

$$\mathrm{H}_\infty(t) = \lim_{n\to\infty} \mathrm{H}_n(t\sqrt{n}) = \exp\left(-\frac{t^2}{2}\right). \tag{10}$$

and holds following property

$$\lim_{\Omega\to\infty} \frac{\ln \mathrm{S}(\Omega)}{\ln \Omega} = -D_2. \tag{11}$$

The correlation dimension can be estimated in this case using model

$$\ln \mathrm{S}(\Omega) = A - D_2 \cdot \ln\ \Omega \tag{12}$$

for parameters A, D_2 using maximum likelihood method and L_4 minimization [11]. Due to smoothness of log likelihood function we obtain asymptotically unbiased point estimate of D_2 and its standard deviation std.

4 Stock Market Indices Analysis

Any stock market index daily closing value can be considered as a realisation of random variable using discrete time. Therefore, there are more ways how to analyse this type of data. The first possible approach is to select several indices and analyse their outcomes as vector in more-dimensional space. Each vector can be again regarded as a realisation of multiple-dimensional random process. The second way of the analysis employs each stock market index separately and estimates the dimension of the trajectory or the dimension of the vectors that can be used for the state space reconstruction. Moreover, due to the very precise estimation of the dimension, one particular index or the whole stock market in general can be investigated within the time. Changes of the dimension between years hold important information about the predictability of the characteristics.

We used 10 stock market indices for the analysis – AEX, DAX, DJI, HSI, NASDAQ, NIKKEI, NYSE, OMX, SMI and SP500. The data contained the daily closing values of the stock markets between 2009-2016. For each index, there is roughly 260 records based on the number of trading days in that year. All experiment in this section were performed using the rotational spectrum (RS) method if not indicated otherwise.

4.1 Dimension Development in Time

The development of the stock market of all ten indices can be investigated by means of the first outlined approach. The dimension development between 2009 and 2016 is illustrated in Figure 1. Whereas the dimension of the stock market in general (all outcomes of the ten selected indices) can vary a lot, the development of correlation in case of particular indexes is more stable. Stock prices were investigated in ten dimensional space first (left top) and the correlation dimension varies between two and three. The same prices analyses were performed for two typical indices (left middle and bottom) as trajectory graphs of correlation dimension between 1 and 2. The same analysis was performed with logarithmic returns as seen in the right column. The rectangles always surround the regions $[D_2 - std; D_2 + std]$, where D_2 is mean value estimate marked with dot and *std* is its standard deviation.

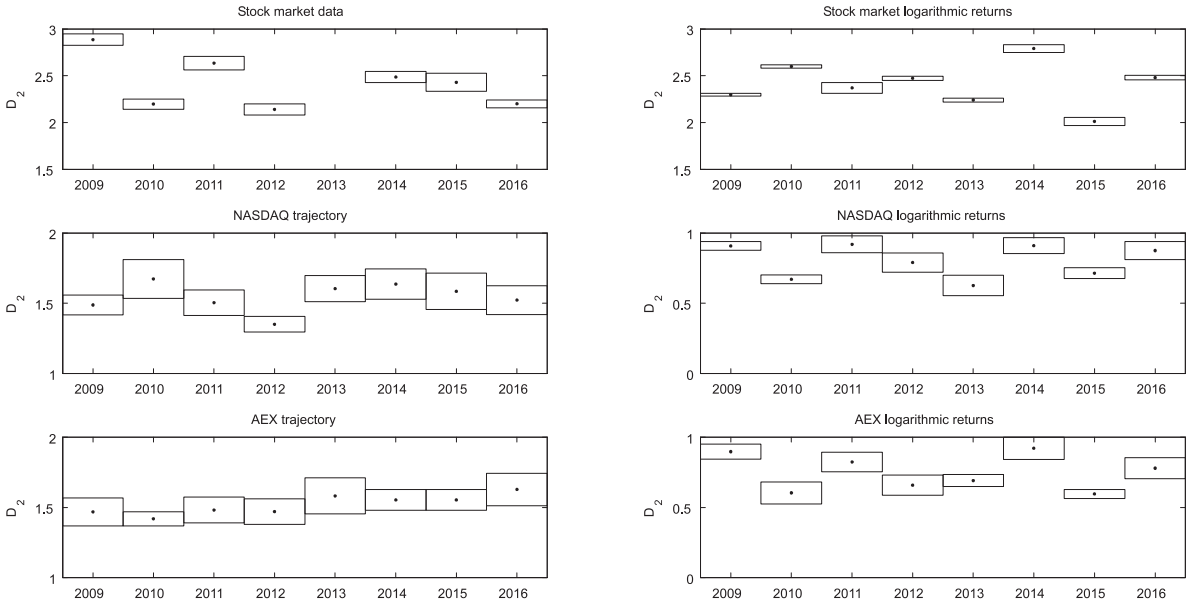


Figure 1 Development of correlation dimensions within the last 8 years as mean values (.) and $\pm std$ boxes.

4.2 Stock Market Behaviour

When analysing the dimension stock market in time, we performed the study in 10-dimensional space. To find out, which indices contributed to the growth of the dimension most, we will divide them for each experiment into two groups, each containing 5 indices. There are in total

$$C = \binom{10}{5} = 252 \text{ combinations.} \tag{13}$$

In this experiment, we will investigate the correlation dimension of the outcomes of each group as a dimension of points in 5-dimensional space. The case when the absolute difference between the dimensions is the biggest is presented in Table 1. The column D_2 depicts the estimate of correlation dimension and std is the standard deviation of the estimate. The dimension difference between these two groups was $\Delta D = 0.8108$.

group	D_2	*std*
AEX, DAX, NIKKEI, OMX, SMI	0.5837	0.0227
DJI, HSI, NASDAQ, NYSE, SP500	1.3945	0.0924

Table 1 Group analysis of stock market indices.

4.3 State Space Reconstruction

Any time series can be investigated by means of the dynamic process and its attractor. We use N-dimensional discrete dynamical process with internal state $\mathbf{s}_k \in \mathbb{R}^n$ and output $x_k \in \mathbb{R}^n$ in time $k \in \mathbb{N}_0$. Using reconstruction length $L \in \mathbb{N}$ we define sliding sample $\mathbf{q}_k = (x_k, \ldots, x_{k+L-1}) \in \mathbb{R}^L$ for $k \in \mathbb{N}_0$, first. Whitney embedding theorem [20] can be rewritten from continuous to discrete time as follows: When $L \geq 2N + 1$, then the reconstructed series $\{\mathbf{q}_k\}_{k=0}^{\infty}$ has the same structure of attractor as the unknown state series $\{\mathbf{s}_k\}_{k=0}^{\infty}$. Therefore, the correlation dimension D_2 of reconstructed attractor is the same as in the case of state series. The only one problem is in the estimation of internal state dimension N. Supposing non-linear autoregressive dependency to other D subjects of depth H, we can estimate internal dimension as $N = (D + 1)H$.

In our experiment, we set for the first experiment $L = 5$ and for the second experiment $L = 10$. The aim was to estimate correlation dimension of the attractor of the original time series. In this case, the logarithmic returns were investigated and the results for different reconstruction lengths L are shown in table 2.

index	D_2	*std*	D_2	*std*
	$L = 5$		$L = 10$	
AEX	0.9694	0.0446	0.9575	0.0282
DAX	0.9448	0.0692	0.9348	0.0228
DJI	0.6943	0.0597	0.7142	0.0322
HSI	0.7562	0.0757	0.7292	0.0349
NASDAQ	0.8205	0.0747	0.8436	0.0388
NIKKEI	0.6957	0.0902	0.7096	0.0561
NYSE	0.6898	0.0638	0.6689	0.0255
OMX	0.8579	0.0755	0.8645	0.0558
SMI	0.3811	0.0600	0.3481	0.0274
SP500	0.7634	0.0764	0.7863	0.0199

Table 2 Attractor correlation dimension.

As easy to see, the correlation dimension estimates for particular index were almost the same in the case of $L = 5$ and in the case of $L = 10$. Therefore the reconstruction length $L = 5$ is sufficient for the fractal analysis of the dynamic process attractor. The highest attractor dimension was captured in the case of AEX stock market and the lowest estimate was observed in the case of SMI index.

4.4 General Dependency of Indices

The following section deals with the estimate of the dimension of the trajectory of the random process whose realisations are the outcomes of the indices. Often, the stock market indices are close to the realisations of Wiener

process, therefore their increments are almost independent. Table 3 presents the dimension estimation using rotational spectrum (RS) and compares it with the correlation sum approach (CD). The rank column sorts the indexes in ascending order based on the D_2 (RS) estimation. All indices except DJI, NASDAQ and OMX have D_2 (RS) < 1.5, and therefore Hurst exponent $H > 1/2$ as evidence of long term memory. The referential CD method is not too accurate and excludes DAX and NYSE with D_2 (CD) > 1.5.

index	D_2 (RS)	*std* (RS)	rank	D_2 (CD)	*std* (CD)
AEX	1.4534	0.0261	3	1.4299	0.1809
DAX	1.4752	0.0485	5	1.5130	0.2322
DJI	1.5188	0.0225	9	1.5899	0.2030
HSI	1.4608	0.0508	4	1.4029	0.1174
NASDAQ	1.5003	0.0398	8	1.5358	0.1825
NIKKEI	1.4410	0.0234	1	1.3989	0.1095
NYSE	1.4808	0.0414	6	1.5264	0.1932
OMX	1.5362	0.0734	10	1.5065	0.1984
SMI	1.4423	0.0591	2	1.4574	0.1393
SP500	1.4963	0.0529	7	1.4984	0.1742

Table 3 D_2 estimation of trajectories of indices.

The standard deviation of the estimate in the case of rotational spectrum is smaller than in the correlation sum approach and will be subject of future research. The mean values and standard deviations are visualised in figure 2. The rectangles surround the regions $[D_2 - \text{std}D_2; D_2 + \text{std}D_2]$ in bold in case of rotational spectrum and with dashed line in the case of correlation sum. The point is the estimate of D_2 using RS and the star is the estimate using CD.

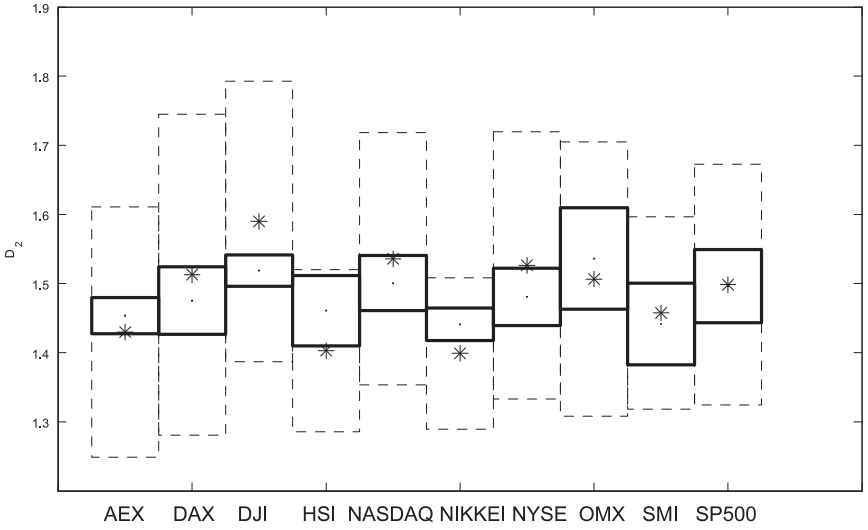


Figure 2 Stock market index dimensions as (.) and solid boxes for RS and (*) and dashed boxes for CD.

5 Conclusion

Numerical experiments showed that stock market indices can be investigated in a variety of ways. In the case of dimension development in time, stock market as a whole (or the 10 most important indices) exhibit significant dimension changes during the last eight years. However, when trying to analyse their trajectories or logarithmic outcomes separately, the dimension change is not that visible. Therefore, the small changes between the predictability of the individual stock market during time can cause big differences in the dependency of the whole stock market. The ten investigated indices were separated into two groups where the first group contained the indices whose outcomes contribute to the dimension of the whole market most and the second group's contributions

are not that substantial. The numerical simulation proved that the changes of the indices AEX, DAX, NIKKEI, OMX and SMI have the biggest effect on the behaviour of the whole stock market.

Further, the state space reconstruction of time series was performed. There was no difference between the analysis using reconstruction length bigger than five elements. The correlation dimension of the attractor of original time series was always in the unit interval and varied from 0.38 to 0.97. The highest attractor dimension was observed in the case of AEX whereas the lowest dimension achieved SMI index. Combined with the previous analysis we conclude that the state space reconstruction cannot provide exact information about the stock market development in general. Focusing on predictability of time series, the trajectories of individual indices were examined by means of rotational spectrum and correlation sum method. The rotational spectrum method gained substantially smaller standard deviation and the indices SMI, NIKKEI and AEX exhibited the smallest changes and therefore they are the most suitable for investments and prediction making.

Acknowledgements

Supported by the grant No. SGS17/196/OHK4/3T/14 of the Czech Technical University.

References

[1] Auer, BR. On time-varying predictability of emerging stock market returns. *Emerging Markets Review*. **27** (2016). 1-13.

[2] Barnett, K. D. On the estimation of the correlation dimension and its application to radar reflection discrimination. *NASA Rep.4564 DOT/FAA/RD-93/41* (1993).

[3] Czarnecka, A. and Gorski, A. How Can the Fractal Geometry Help with Analyze of the Stock Exchange Market? *Information Systems Architecture and Technology - ISAT 2016, PT IV*. **524** (2017) 45-53.

[4] Debye, P. Zerstreuung von Rontgenstrahlen. *Annalen der Physik* **6** (1915).809-823.

[5] Dhifaoui, Z. Robust to noise and outliers estimator of correlation dimension. *Chaos, Solitons & Fractals*. **93** (2016) 169-174.

[6] Dlask, M. Fractional Brownian Bridge as a Tool for Short Time Series Analysis. In: *34th International Conference Mathematical Methods in Economics (MME)*, 2016, 149154.

[7] Dlask M. and Kukal J. Application of Rotational Spectrum for Correlation Dimension Estimation. *Chaos, Solitons & Fractals/* **99** (2017) 256-262.

[8] Dlask, M. and Kukal J. Revisited Zero-Crossing Method for Hurst Exponent Estimation in Time Series. In: *33th International Conference Mathematical Methods in Economics (MME)*, 2015, 115120.

[9] G. Edgar, Measure, Topology, and Fractal Geometry (Undergraduate Texts in Mathematics), Springer, 2013.

[10] K. J. Falconer, Fractal geometry : mathematical foundations and applications, Wiley, Chichester, 2003.

[11] Ge D., Jiang X. and Ye Y. A note on the complexity of Lp minimization. *Mathematical Programming* **129** (2011), 285-299.

[12] Grassberger, P., Procaccia, I. Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena* **9** (1983), 189-208.

[13] Grassberger, P., Procaccia, I. Characterization of Strange Attractors. *Physical Review Letters* **5** (1983), 346-349.

[14] Jiang, JQ. Using Renyi parameter to improve the predictive power of singular value decomposition entropy on stock market. *Physica A - Statistical Mechanics and Its Applications*. **448** (2016), 254-264

[15] Kalauzi A., Vuckovic A. and Bojic T., Topographic distribution of EEG alpha attractor correlation dimension values in wake and drowsy states in humans. *International Journal of Psychophysiology 95* **3** (2015), 278291.

[16] Krakovska, A. Correlation dimension underestimation. *Acta physica slovaca* **5** (1995), 567-574.

[17] Mandelbrot, B. B. and Van Ness, J. W. Fractional Brownian motions,Fractional Noises and Applications. *SIAM Review* **10** (1968), 422-437.

[18] Nie, CX. Correlation dimension of financial market. *Physica A - Statistical Mechanics and Its Applications.* **473** (2017), 632-639.

[19] Shayegh F.,Sadri S. and Amirfattahi R. A model based method for computation of correlation dimension, lyapunov exponents and synchronization from depth-EEG signals. *Computer Methods and Programs in Biomedicine 113* **1**(2014), 323337.

[20] Whitney, H. Differentiable Manifolds. *Annals of Mathematics* **37** (1936), 648-80.

[21] Zhou, S. Feng, Y. Wu, WY. A novel method to identify the scaling region of correlation dimension. *Acta Physica Sinica* **64** (2015).

A DEA-Based Inequality Measure: Application to Allocation of Health Resources

Martin Dlouhý[1]

Abstract. Data envelopment analysis (DEA) is a non-parametric method of production function estimation, which is able to deal with multiple inputs and outputs. The objective of this study is to show how a resource-allocation DEA model is able to measure the inequality in case of multiple resources. The DEA model will be applied to the health systems because inequalities in the geographic allocation of health resources are an important policy issue even in developed European countries. Most comparisons of health data in Europe take place at the national level. However, there is an increased interest in looking at health data at a sub-national level. For this purpose, we use the data that come from the regional statistics of the Eurostat and the unit of analysis is the NUTS 2 region. The inequality measure based on DEA is calculated for seven selected European countries.

Keywords: Data Envelopment Analysis; Resource Allocation; Inequality; Regional Statistics; NUTS 2.

JEL Classification: I14, C60
AMS Classification: 90C05, 91B15

1 Introduction

1.1 Inequality measurement

Measures of inequality express the complexity of variation in observed variable by a single number. The simple measures of inequality are the ranges, which use only data on the extreme values. The absolute range is defined as a difference between the maximum and minimum observed values per capita ($\gamma_{max} - \gamma_{min}$). The relative range is defined as $(\gamma_{max} - \gamma_{min})/\gamma_{pop}$, where γ_{pop} is the average number of units per capita for the entire population. Other measure of this type is the maximin ratio, which is defined as $\gamma_{max}/\gamma_{min}$. By concentrating on the geographical areas with extreme values only, all these indices give only a limited view on the overall distribution and fail to possess many desirable properties of inequality indices required by theory.

The widely known measure of inequality that uses all observations is the Gini coefficient. The Gini coefficient is derived from the Lorenz curve, a cumulative frequency curve that compares the empirical distribution of the studied variable with the uniform (egalitarian) distribution that represents the perfect equality. The Gini coefficient ranges between 0, which occurs in case of the perfect equality, and 1, which occurs in case of the perfect inequality. The Robin Hood Index measures what proportion of resources has to be moved from areas with above-average provision to areas with below-average provision to achieve equal distribution. The main advantage of the Robin Hood Index over the Gini coefficient and other inequality measures, such as the Atkinson index and Theil's entropy measure, is its clear interpretation. The Robin Hood Index (RHI) is calculated by the formula:

$$RHI = \frac{1}{2}\sum_{i=1}^{n}\left|\pi_i - \rho_i\right|, \tag{1}$$

where π_i is the population proportion, ρ_i is the resource proportion, and n is the number of geographical areas. The RHI is usually multiplied by 100 to be in percentages.

Does the choice of inequality measure matter? Under some circumstances it could; however, there is some evidence that the most common indicators are usually interchangeable. For example, Kawachi and Kennedy [13] calculated the income distribution for the 50 U.S. states and studied the relation of income inequality to mortality. Kawachi and Kennedy examined the following measures of income distribution: the Gini coefficient, the decile ratio, the proportions of total income earned by bottom 50%, 60%, and 70% of households, the Robin Hood Index, the Atkinson Index, and Theil's entropy measure. All inequality measures were highly correlated with each other, and in no instance did the correlation coefficient fall below 0.86 in the absolute value. Kawachi and Kennedy [13] concluded that a theoretical justification for the choice of indicator is critical in assessing the impact of social and

[1] University of Economics, Prague, 4 Winston Churchill Sq., Prague 3, Czech Republic, dlouhy@vse.cz.

economic policies on income distribution and mortality. However, there is a little evidence to suggest that the choice of indicator will result in an absolutely different conclusion.

1.2 Inequality in Health Care

The free market allocates health resources according to willingness and ability to pay, not according to the health needs of local population. This means that a supply of health services will be concentrated in rich areas, whereas poor areas, albeit being usually those with greatest health needs, will not be served adequately. However, one should take into account that the European health systems are mostly publicly funded and are highly regulated, so the unequal distribution can be a likely consequence of wrong public regulation than a free market. The geographic inequalities in many European countries are clearly the results of state planning, so the efficient use of resources in public sector can be questioned. In the public health system, the equal access to services is one of the main objectives of national health policy. Therefore, any variation in the distribution of health resources may be considered as an inequality whether it be a consequence of free market or wrong public regulation.

The existence of significant differences in health resource inputs, health expenditures, the utilization of health services, and health outcomes have been documented by a lot of studies from various countries of the world. For example, Johnston and Wilkinson [12] studied the distribution of general practitioners in Australia between 1986 and 1996. They used crude mortality as a measure of community need for medical services. The Robin Hood Index was used as a measure of overall distribution. Nationally, the number of people sharing each general practitioner fell by 11% from 1,038 in 1986 to 921 in 1996. However in 41 of 57 areas the number of people per general practitioner actually increased over this period, indicating increasing inequity in the distribution. Over the decade, the number of relatively under-served areas increased from 67% in 1986 to 79% in 1996. Thus, despite the increasing number of general practitioners overall, the rural and remote parts of Australia became increasingly poorly served.

Dlouhý [6] analysed the regional distribution of outpatient services in the Czech Republic between 1996 and 2002. The supply of outpatient services was measured by the number of outpatient physicians in full-time equivalents in both independent practices and outpatient hospital departments. As a measure of health need, the population of the region was used. The Gini coefficient and the Robin Hood Index indicated that the overall distribution of physicians in outpatient services was relatively stable over the period 1996-2002. Total numbers of physicians in outpatient services may hide inequalities within the medical specialties though. For example in psychiatry, there were 12.63 psychiatrists per 100,000 inhabitants in Prague, but only 3.24 in the Region of Zlín. The differences of that magnitude could not be explained by the differences in the health needs, but are rather results of wrong regulatory policies.

The study of OECD [15] presents information on geographic variations in health care utilisation within and across 13 OECD countries. The analysis focuses on a selected set of high-volume and high-cost health care activities. Health care utilisation is recorded at the patient's place of residence. Hence, the level of use in a given area cannot be explained by patients receiving treatment in other geographic areas. While the analysis in this study does not allow to determine precisely how much of these variations are unwarranted, some of these variations are too large to be explained solely by patient needs and/or preferences [15].

Dlouhý [7] evaluated the inequalities in the geographic distribution of doctors and hospital beds in the European countries. The unit of analysis is the NUTS 2 region. In total, 13 countries were included in the sample (Austria, Bulgaria, Czech Republic, France, Hungary, Italy, Norway, Poland, Portugal, Romania, Slovakia, Spain, and Turkey). The Robin Hood Index was calculated for selected countries to measure regional inequalities in the distribution of doctors and hospital beds among the NUTS 2 regions. In 10 of 13 countries, the differences in regional distribution are higher for doctors than those for hospital beds. Three highest regional inequalities in case of doctors were found in Slovakia, Hungary, and Turkey. In case of hospital beds, the highest inequalities were identified in Portugal, Spain, and Poland.

The objective of this study is to show how a resource-allocation DEA model is able to measure the inequalities in case of multiple resources. The method is applied to inequalities in the distribution of health resources.

2 Methods and Data

2.1 Data Envelopment Analysis

Data envelopment analysis (DEA) is a method based on the mathematical programming and was originally developed to construct production frontier and evaluate technical efficiency of production units. DEA specifies the production frontier as the piecewise linear envelopment of the data and constructs efficiency measures based on

radial uniform contractions or expansions from inefficient observations to the frontier. The DEA model for multiple inputs and outputs was formulated and solved by Charnes, Cooper, and Rhodes in 1978 [3]. Since 1978 a great variety of DEA models with various extensions and modifications has been developed. These extensions can be found, for example, in textbooks such as Charnes, Cooper, Lewin, and Seiford [2], Cooper, Seiford, and Tone [4], Charnes, Cooper, Zhu [5], Jablonský and Dlouhý [11]. These textbooks also present many examples of applications from both private and public sectors.

Each production unit allocates its resources into a number of inputs to produce various outputs. DEA uses quantities of inputs consumed and outputs produced to calculate the relative technical efficiency of a production unit. The relative technical efficiency of the unit is defined as the ratio of its total weighted output to its total weighted input or, vice versa, as the ratio of its total weighted input to its total weighted output. DEA allows each production unit to choose its own weights of inputs and outputs in order to maximize its efficiency score. A technically efficient production unit is able to find such weights that it lies on the production frontier. The production frontier represents the maximum amounts of output that can be produced by given amounts of input (in the output maximization model) or, alternatively, the minimum amounts of inputs required to produce the given amount of output (in the input minimization model).

For each production unit, the DEA model calculates the efficiency score; determines the relative weights of inputs and outputs; and identifies peers for each production unit that is not technically efficient. The peers of a technically inefficient production unit are technically efficient production units with similar combinations of inputs and outputs. The peers serve as benchmarks, which show potential improvements that the technically inefficient production unit can attain. Because the peers are real production units, one can expect that the efficiency improvements should be attainable by the inefficient units.

A construction of a production frontier and calculation of the efficiency score by the constant returns-to-scale DEA model are shown in Figure 1. Let us suppose there are three production units *A*, *B*, and *C*. These production units produce the same level of single output with two inputs *A* = (10, 10), *B* = (30, 5), and *C* = (20, 10). The production units *A* and *B* are technically efficient and they both lie on the production frontier. The production unit *C* uses more inputs than it is technically necessary; therefore, this unit lies above the production frontier and is technically inefficient. The units *A* and *B* are the peers showing to unit *C* how to reduce both inputs to be technically efficient. The hypothetical production unit *C** is a linear combination of real units *A* and *B*. The hypothetical production unit *C** = (16.67, 8.33) shows the alternative in which all inputs of original unit *C* are reduced proportionately. However, it should be noticed that unit *C* can achieve technical efficiency by moving to any position on the production frontier.

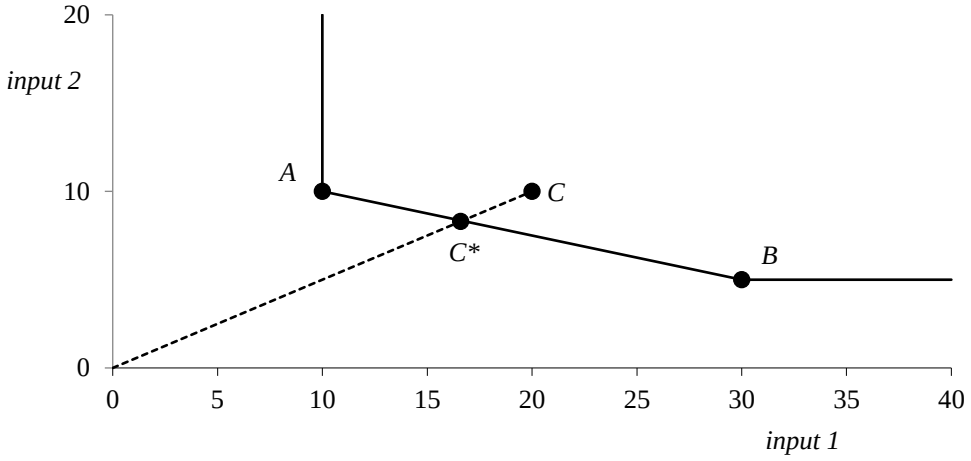


Figure 1 Efficiency Frontier (2 inputs, 1 output)

The so-called CCR DEA model assumes constant returns to scale. Let us have the set of *n* production units that use *m* inputs to produce *r* outputs. The mathematical formulation of the output-oriented version of the CCR DEA model [3] for production unit *q* is:

$$
\begin{aligned}
&\textit{Minimize} && \phi_q = \sum_{i=1}^{m} v_i x_{iq}, \\
&\textit{subject to} && \sum_{k=1}^{r} u_k y_{kj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1, 2, \ldots, n, \\
& && \sum_{k=1}^{r} u_k y_{kq} = 1, \\
& && u_k \ge \varepsilon, \quad i = 1, 2, \ldots, r, \\
& && v_i \ge \varepsilon, \quad j = 1, 2, \ldots, m.
\end{aligned} \tag{2}
$$

where φ_q is the efficiency score, $\boldsymbol{x_{ij}}$ is the amount of input i used by production unit j, y_{kj} is the amount of output k produced by production unit j, and ε is infinitesimal constant. The output weights u_i and input weights v_j are variables in the DEA model. In the output-oriented model, the efficiency score φ_q is one if the unit q is technically efficient, and is greater than one if the unit is technical inefficient. The efficiency score measures a size of output expansion that makes unit q technically efficient. In the input-oriented model, the efficiency score is one if the unit q is technically efficient, and is lower than one if the unit is technical inefficient The obtain efficiency scores for all units, the model (2) has to be solved for each production unit q = 1, 2, ..., n.

The DEA models originally assume input and output independence. However, in some cases, total input or total output is fixed. Let us suppose one output that represent a total budget *TB*, which is fixed and should be distributed by the higher authority among n units based on their inputs. If some money are additionally allocated to one unit, then budgets of other units have to be reduced. That is why this model is called as the Zero sum gains DEA model (ZSG DEA). In the ZSG DEA model, the described resource allocation problem is solved by making such allocation that all units become technically efficient [1, 9, 10, 14]. In the first step, the amount *TB*/n is allocated to each unit and we solve an output-oriented DEA model with the single constant output and original inputs. In the second step, each unit obtains individual budget, which is calculated as $(\varphi_q / \Sigma\varphi_q)TB$. All units are now technically efficient, which means that total budget was distributed efficiently.

2.2 The DEA-Based Inequality Measure

Let us suppose that we want to measure inequality in geographical distribution in case of multiple resources. By using traditional measures introduced in section 1.1 separately for each resource, we will miss the possibility of substitution between resources. For example, the health resources as doctors and nurses are, at least to some extent, substitutes. Hence the region with fewer doctors may compensate such disadvantage by larger number of nurses. To cope with cases with multiple resources, one can use multiple criteria decision making (MCDM) and set the nationwide resource weights. More flexible approaches to estimate resource substitution are econometric modelling and DEA. In this paper, we will concentrate on DEA.

Let us have a situation with two resources (inputs) and one output that can be illustrated by Figure 1. Production units *A*, *B*, and *C* are now geographical areas (regions), hence the output is the regional population that serves as an estimation of need. According to the DEA methodology, regions *A* and *B* that lie on the production frontier are technically efficient, and region *C* is inefficient with the output-oriented efficiency score φ_3 = 1.2. A higher level of inefficiency in this situation represents a higher level of resources that are available for regional population. The efficiency score of the output-oriented model, which is greater than one, expresses the excess of resources above the most badly served regions that are represented by the set of DEA efficient units. The efficiency score φ_q = 1.2 means that the given combination of resources is able to serve a 20% larger population or that there is a 20% excess of resources in the given region. By the efficiency score, multiple resources are now summarized to a single number that can enter the Robin Hood Index calculation.

The output-oriented ZSG DEA model with two inputs (doctors and hospital beds) and one output (regional population) now calculates efficiency scores φ_q that are used to estimate resource-efficient levels of regional populations. The total population of the country, which is naturally a fixed output, is allocated to all regions so that they become technically efficient with the given levels of resources. In other words it means equal distribution of resources.

The result of the ZSG DEA model can be used for the calculation of the Robin Hood Index by formula (1), where π_i is the real population proportion, and ρ_i is the population proportion estimated by the ZSG DEA model.

Thus as inequality measure uses the percentage of the country population that has to move from highly technically efficient regions to technically inefficient regions to achieve technical efficiency for all regions.

2.3 Data

The chosen unit of analysis is the NUTS 2 region. The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the EU. The NUTS classification is defined only for the Member States of the EU. Eurostat, in agreement with the countries concerned, also defines a coding of Statistical Regions for countries that do not belong to the EU but are either candidate countries awaiting accession to the EU or countries belonging to the European Free Trade Association. The NUTS 2013 classification is valid from 1 January 2015 and lists 98 regions at NUTS 1, 276 regions at NUTS 2 and 1342 regions at NUTS 3 level.

All data come from 2013 and were obtained from the Eurostat regional statistics [8]. The sample include seven countries: Austria. Bulgaria, the Czech Republic, France, Poland, Spain, and Turkey). To avoid outliers, we excluded the French overseas territories and Spanish regions of Ciudad Autónoma de Ceuta and Ciudad Autónoma de Melilla. Further, we joined together data on two NUTS 2 regions in cases if the region of the capital was located inside another region. We assume that the population of such neighbouring regions use frequently health services in the capital. This assumption was used in two cases: Prague and Central Bohemia (Czech Republic), Wien and Niederösterreich (Austria).

3 Application

The basic country characteristics are in Table 1. In this study we will consider two health resources: doctors and hospital beds. The highest relative level of resources is found in Austria. On the contrary, the lowest level of health resources is in Turkey. We assume that substitution between doctors and hospital beds is possible. In a region with more doctors, the intensity of care is higher, so a lower number of hospital beds is needed and vice versa.

Country	Population	Number of Regions (adjusted)	Doctors per 1000	Hospital Beds per 1000
Austria	8 451 860	8	5.01	7.67
Bulgaria	7 284 552	6	3.77	6.29
Czech Republic	10 516 125	7	3.69	6.46
France	63 697 865	22	3.35	6.34
Poland	38 062 535	16	2.24	6.58
Spain	46 559 731	17	3.81	2.96
Turkey	75 627 384	26	1.77	2.67

Table 1 Basic Country Characteristics

The method is applied to seven countries described in the data section. The number of regions (units in the DEA evaluation) for each country was already defined in Table 1. First, the ZSG DEA model was calculated. Secondly, the Robin Hood Index was calculated as a difference between real population proportion π_i and predicted population proportion ρ_i obtained from the ZSG DEA model. For a comparison, the RHIs were calculated separately for doctors and hospital beds by the traditional approach. As expected, the DEA-based RHIs showed lower inequality than separate RHIs (Table 2).

Country	Robin Hood Index Doctors	Robin Hood Index Hospital beds	Robin Hood Index ZSG-DEA
Austria	6.46	4.32	3.47
Bulgaria	4.16	4.07	3.94
Czech Republic	6.38	2.45	2.45
France	5.99	3.45	3.28
Poland	5.45	8.16	4.39
Spain	8.46	8.68	8.46
Turkey	9.55	7.71	7.71

Table 2 Robin Hood Index

4 Conclusion

The DEA-based inequality measure was formulated on the basis of resource-allocation ZSG DEA model. The inequality measure was calculated for seven European countries to measure regional inequalities in the distribution of doctors and hospital beds. In all seven countries, the DEA-based Robin Hood Index was the same or lower level of inequality. There are two areas of further research: first, the possible use of other DEA models for inequality measurement; second, formulations of the DEA-based versions of other inequality measures.

Acknowledgements

The research was supported by the grant no. 16-01821S of the Czech Science Foundation.

References

[1] Bi, G., Feng, C., Ding, J., Liang, L., and Chu, F.: The linear formulation of the ZSG-DEA models with different production technologies. *Journal of the Operational Research Society* **65** (2014), 1202–1211.
[2] Charnes, A., Cooper, W. W., Lewin, A. Y, and Seiford, L. M.: *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer Academic Publishers, Boston, 1994.
[3] Charnes, A., Cooper, W. W., and Rhodes, E.: Measuring the inefficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.
[4] Cooper, W. W., Seiford, L.W., and Tone, K.: *Data envelopment analysis*. Kluwer Academic Publishers, Boston, 2000.
[5] Cooper, W. W., Seiford, L.W., and Zhu, J.: *Handbook of Data Envelopment Analysis*. Kluwer Academic Publishers, Boston, 2004.
[6] Dlouhý, M.: Health care regulation: impact on the supply of outpatient services in the Czech Republic. *Prague Economic Papers* **14** (2005), 3:267–276.
[7] Dlouhý, M.: Inequalities in geographic distribution of doctors and hospital beds in 13 European countries. In: *Proceedings of the 21st International Conference Current Trends in Public Sector Research 2017*. Masaryk University, Brno, 2017, pp. 254–259.
[8] Eurostat: *Regional statistics by NUTS classification*. European Commission, 2016. Available on http://ec.europa.eu/eurostat/web/regions/data/database.
[9] Gomes, E. G., and Lins, M. P. E.: Modelling undesirable outputs with zero sum gains data envelopment analysis models. *Journal of the Operational Research Society* **59** (2008), 616–623.
[10] Gomes, E. G., and Da Silva e Souza, G.: Allocating financial resources for competitive projects using a zero sum gains DEA model. *Engevista* **12** (2010), 4–9.
[11] Jablonský, J., and Dlouhý, M.: *Modely hodnocení efektivnosti a alokace zdrojů* [Models of Efficiency Evaluation and Resource Allocation]. Professional Publishing, Prague, 2015.
[12] Johnston, G., and Wilkinson, D.: Increasingly inequitable distribution of general practitioners in Australia, 1986–96. *Australian and New Zealand Journal of Public Health* **25** (2001), 1:66–70.
[13] Kawachi, I., and Kennedy, B. P.: The relationship of income inequality to mortality: does the choice of indicator matter? *Social Science and Medicine* **45** (1997), 7:1121–1127.
[14] Lins, M. P. E., Gomes, E. G., Soares De Mello, J. C. C. B., and Soares De Mello, A. J. R.: Olympic ranking based on a zero sum gains DEA model. *European Journal of Operational Research* **148** (2003), 312-322.
[15] OECD. *Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health System Performance?* OECD Health Policy Studies, OECD Publishing, 2014.

Solvability of interval max-plus matrix equations

Emília Draženská [1] [2]

Abstract. Max-plus algebra is an algebraic structure in which the usual operations of multiplication and addition of two numbers are replaced by the operations arithmetical addition and selection of the greater of the two numbers, respectively. Many situations of daily life can be described by max-plus equations. Especially, equation in which intervals can be substituted for the exact numbers, because it is often useful. So, we assume such a matrices, in which numbers will be replaced by intervals, it is called an interval matrices. And we will study solvability of interval matrix equations. There are several types of solvability of interval max-plus matrix equation. In this paper, we give necessary and sufficient conditions for four of them.

Keywords: max-plus algebra, interval matrix eguations, solvability

JEL classification: C02
AMS classification: 15A18; 15A80; 65G30

1 Motivation

Many discrete event dynamic systems can be described using max-plus algebra. Max-plus algebra is an algebraic structure in which the usual operations of multiplication and addition of two numbers are replaced by addition and maximum, respectively.

Results of the solvability of max-plus linear equations $A \otimes x = b$, where A is a matrix, b and x are vectors, appeared several years ago. Systems of max-plus linear equations are used in several branches of applied mathematics.

In this paper, we study two-sided matrix equations in the form $A \otimes X \otimes C = B$, where A, B, and C are given matrices and X is an unknown matrix. In following example, we give one of possible application of such a matrix equation.

Example 1. Consider the manufacturing company that carries its three types of products in three places P_1, P_2, and P_3. Every product is placed in two warehouses W_1 and W_2. Afterwards, they are loaded into two trucks T_1 and T_2 which deliver products to three shops S_1, S_2 and S_3 (see Figure 1).

In Figure 1, there is an arrow from P_i to W_j if the products made in the place P_i are stored in warehouse W_j. If products from warehouse W_j are loaded on a truck T_k, then there is an arrow from W_j to T_k. And, there is an arrow from T_k to S_l, if the truck T_k expedites a products to the shop S_l ($i = 1, 2, 3$; $j = 1, 2$; $k = 1, 2$; $l = 1, 2, 3$).

The symbols above or below the arrows, a_{ij} (c_{kl}), express the time needed to transport goods from place P_i to warehouse W_j (from truck T_k to shop S_l). The time needed to transport goods from warehouse W_j to truck T_k is denoted by x_{jk}. Then, total time needed to transport goods from place P_i to shop S_l is $a_{ij} + x_{jk} + c_{kl}$.

Let us denoted by b_{il} the time needed to transport from place P_i to shop S_l To arrange the transportation for all products made in P_1 to the shops, the following equations must be satisfied:

$$\begin{aligned} &\max\{a_{11} + x_{11} + c_{11},\ a_{11} + x_{12} + c_{21},\ a_{12} + x_{21} + c_{11},\ a_{12} + x_{22} + c_{21}\} = b_{11}. \\ &\max\{a_{11} + x_{11} + c_{12},\ a_{12} + x_{21} + c_{12}\} = b_{12}, \\ &\max\{a_{11} + x_{12} + c_{23},\ a_{12} + x_{22} + c_{23}\} = b_{13}, \end{aligned} \tag{1}$$

Similar equalities must be satisfied to arrange the transportation for all products from P_2 and P_3 to the shops $S_1,\ S_2$ and S_3.

In general, let us suppose that there are m places $P_1, P_2, \ldots, P_m$, n warehouses $W_1, W_2, \ldots, W_n$, s trucks $T_1, T_2, \ldots, T_s$, and r shops $S_1, S_2, \ldots, S_r$. If there is no connection from P_i to W_j (from T_k to S_l), we put $a_{ij} = -\infty$ ($c_{kl} = -\infty$). Let us introduce the following notations: sets $M = \{1, 2, \ldots, m\}$, $N = \{1, 2, \ldots, n\}$, $R = \{1, 2, \ldots, r\}$, and $S = \{1, 2, \ldots, s\}$.

So, we would like to determine times x_{jk} for any $j \in N$ and for any $k \in S$ such that the maximum of total transport time from P_i to S_l is equal to a given number b_{il} for any $i \in M$ and for any $l \in R$. Thus,

$$\max_{j \in N,\, l \in S} \{a_{ij} + x_{jk} + c_{kl}\} = b_{il}. \tag{2}$$

[1] The research was supported by the Slovak VEGA grant No. 1/0389/15.

[2] Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovak Republic, emilia.drazenska@tuke.sk

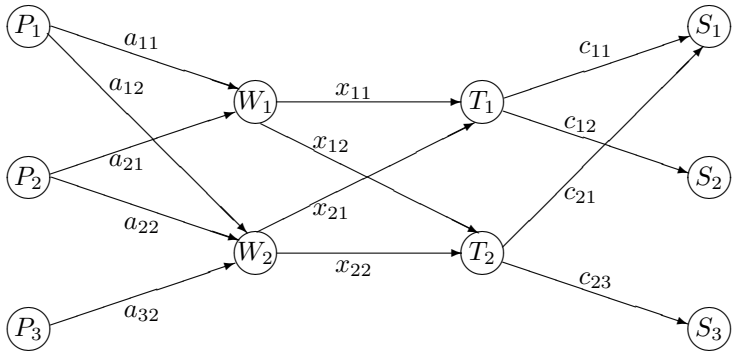


Figure 1 Transport diagram

2 Max-plus algebra

Max-plus algebra is the triple $(\overline{\mathbb{R}}, \oplus, \otimes)$, where

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\},\ a \oplus b = \max\{a, b\} \text{ and } a \otimes b = a + b.$$

The set of all $m \times n$ *matrices* over $\overline{\mathbb{R}}$ is denoted by $\overline{\mathbb{R}}(m, n)$ and the set of all column n-*vectors* over $\overline{\mathbb{R}}$ by $\overline{\mathbb{R}}(n)$.
Operations $\oplus$ and $\otimes$ are extended to matrices and vectors in the same way as in the classical algebra. We will consider the *ordering* $\leq$ on the sets $\overline{\mathbb{R}}(m, n)$ and $\overline{\mathbb{R}}(n)$ defined as follows:

- for any $A, C \in \overline{\mathbb{R}}(m, n):\ A \leq C$ if and only if $a_{ij} \leq c_{ij}$ for each $i \in M$ and for each $j \in N$,
- for any $x, y \in \overline{\mathbb{R}}(n):\ x \leq y$ if and only if $x_j \leq y_j$ for each $j \in N$.

We will use the *monotonicity of* $\otimes$. That means:
for each $A, C \in \overline{\mathbb{R}}(m, n)$ and $B,\ D \in \overline{\mathbb{R}}(n, s)$: if $A \leq C$ and $B \leq D$, then $A \otimes B \leq C \otimes D$.

3 Interval matrix equations

Let $A \in \overline{\mathbb{R}}(m, n)$, $B \in \overline{\mathbb{R}}(m, r)$ and $C \in \overline{\mathbb{R}}(m, n)$ be matrices with elements a_{ij}, b_{ik}, and c_{lk}, respectively. We rewrite the system of equalities (2) from Example 1 in the matrix form

$$A \otimes X \otimes C = B \tag{3}$$

Based on the real situation, often, the elements of given matrices in the system of equations are not numbers, but intervals of numbers. So, in this paper, we will deal with the solvability of interval max-plus matrix equations.

Similarly to [1], [2], [5], [6] we define *interval matrices* $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}, \boldsymbol{X}$ in following way:

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \overline{\mathbb{R}}(m, n);\ \underline{A} \leq A \leq \overline{A} \right\}, \quad \boldsymbol{B} = [\underline{B}, \overline{B}] = \left\{ B \in \overline{\mathbb{R}}(m, r);\ \underline{B} \leq B \leq \overline{B} \right\},$$

$$\boldsymbol{C} = [\underline{C}, \overline{C}] = \left\{ C \in \overline{\mathbb{R}}(s, r);\ \underline{C} \leq C \leq \overline{C} \right\}, \quad \boldsymbol{X} = [\underline{X}, \overline{X}] = \left\{ X \in \overline{\mathbb{R}}(n, s);\ \underline{X} \leq X \leq \overline{X} \right\}.$$

We will write the notation

$$\boldsymbol{A} \otimes \boldsymbol{X} \otimes \boldsymbol{C} = \boldsymbol{B} \tag{4}$$

for the set of all matrix equations of the form (3) such that $A \in \boldsymbol{A},\ B \in \boldsymbol{B}$, and $C \in \boldsymbol{C}$ and each solution is required from the set $\boldsymbol{X}$. We call (5) an *interval max-plus matrix equation*.

A special case of (4) is

$$\boldsymbol{A} \otimes X \otimes \boldsymbol{C} = \boldsymbol{B}, \tag{5}$$

in which $X \in \overline{\mathbb{R}}(n, s)$.

4 T3, T8, T9, and T10 Solvability

In [4] and in [3], there were studied several solvability concepts of (5). In this paper, we will deal with the solvability of (4). We define some types of solvability.

Definition 1. Interval max-plus matrix equation in the form (4) is
- *T10 solvable* if and only if for each $X \in \overline{\mathbb{R}}(n,s)$, for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$ the product $A \otimes X \otimes C$ lies in $\boldsymbol{B}$;
- *T3 solvable* if and only if for each $X \in \overline{\mathbb{R}}(n,s)$ there exists $B \in \boldsymbol{B}$ such that for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$ holds $A \otimes X \otimes C = B$.
- *T8 solvable* if and only if for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$ there exist $B \in \boldsymbol{B}$ such that for each $X \in \overline{\mathbb{R}}(n,s)$ holds $A \otimes X \otimes C = B$.
- *T9 solvable* if and only if there exist $B \in \boldsymbol{B}$ such that for each $A \in \boldsymbol{A}$, for each $C \in \boldsymbol{C}$ and for each $X \in \overline{\mathbb{R}}(n,s)$ holds $A \otimes X \otimes C = B$.

Theorem 1. *Interval max-plus matrix equation* (4) *is T10 solvable if and only if the system of inequalities*

$$\underline{A} \otimes \underline{X} \otimes \underline{C} \geq \underline{B}, \tag{6}$$

and

$$\overline{A} \otimes \overline{X} \otimes \overline{C} \leq \overline{B} \tag{7}$$

are satisfied.

Proof. T10 solvability means that the system of inequalities $\underline{B} \leq A \otimes X \otimes C \leq \overline{B}$ holds for each $A \in \boldsymbol{A}$, $C \in \boldsymbol{C}$ and for each $X \in \boldsymbol{X}$. The left inequality holds for each $A \in \boldsymbol{A}$, $C \in \boldsymbol{C}$ and for each $X \in \boldsymbol{X}$ if and only if it holds for $A = \underline{A}$, $C = \underline{C}$ and $X = \underline{X}$. Thus, the inequality (6) is true. Similarly, the validity of the right inequality is equivalent to (7). □

For a given $u \in M,\ v \in N$ denote the matrix $A^{(uv)} = (a_{ij}^{(uv)})$, where

$$(a_{ij}^{(uv)}) = \begin{cases} \overline{a}_{ij} & \text{for } i = u,\ j = v, \\ \underline{a}_{ij} & \text{otherwise.} \end{cases}$$

Lemma 1. Let $A \in \boldsymbol{A}$. Then

$$A = \bigoplus_{u \in N, v \in S} (a_{uv} - \overline{a}_{uv}) \otimes A^{(uv)}. \tag{8}$$

Proof. Let $i \in M,\ j \in N$. Then

$$\bigoplus_{u \in N, v \in S} (a_{uv} - \overline{a}_{uv}) \otimes A^{(uv)} = [\bigoplus_{u \in N, v \in S} (a_{uv} - \overline{a}_{uv}) \otimes A^{(uv)}]_{ij} = \bigoplus_{u \in N, v \in S} (a_{uv} - \overline{a}_{uv}) \otimes a_{ij}^{(uv)} =$$

$$= \big(\bigoplus_{(u,v) \neq (i,j)} (a_{uv} - \overline{a}_{uv}) \otimes a_{ij}^{(uv)}\big) \oplus ((a_{ij} - \overline{a}_{ij}) \otimes a_{ij}^{(ij)}) = \big(\bigoplus_{(u,v) \neq (i,j)} (a_{uv} - \overline{a}_{uv}) \otimes \underline{a}_{ij}\big) \oplus ((a_{ij} - \overline{a}_{ij}) \otimes \overline{a}_{ij}) =$$

$$= \big(\bigoplus_{(u,v) \neq (i,j)} (a_{uv} - \overline{a}_{uv} + \underline{a}_{ij})\big) \oplus (a_{ij} - \overline{a}_{ij} + \overline{a}_{ij}) = \big(\bigoplus_{(u,v) \neq (i,j)} (a_{uv} - \overline{a}_{uv} + \underline{a}_{ij})\big) \oplus a_{ij} = a_{ij},$$

because $a_{uv} - \overline{a}_{uv} + \underline{a}_{ij} \leq \underline{a}_{ij} \leq a_{ij}$. So, the equality (8) is satisfied. □

It is easy to see, using definitions of the T3, T10 and T8 solvability, that the T10 solvability is a necessary condition for T3 solvability and also for T8 solvability.

Lemma 2. Interval max-plus matrix equation (4) is T3 solvable if and only if interval max-plus matrix equation (4) is T10 solvable and the equality

$$\underline{A} \otimes X \otimes \underline{C} = \overline{A} \otimes X \otimes \overline{C} \tag{9}$$

is satisfied for each $X \in \boldsymbol{X}$.

Proof. The T10 solvability implies that $\underline{A} \otimes X \otimes \underline{C} \geq \underline{A} \otimes \underline{X} \otimes \underline{C} \geq \underline{B}$ and $\overline{A} \otimes X \otimes \overline{C} \leq \overline{A} \otimes \overline{X} \otimes \overline{C} \leq \overline{B}$. It means that $\underline{A} \otimes X \otimes \underline{C} \in \boldsymbol{B}$ and $\overline{A} \otimes X \otimes \overline{C} \in \boldsymbol{B}$ for each $X \in \boldsymbol{X}$. Let $X \in \boldsymbol{X}$ be fixed. The existence of $B \in \boldsymbol{B}$ such that for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$ the equality $A \otimes X \otimes C = B$ is satisfied means that the products $A \otimes X \otimes C$ are equal to each other for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$, which is equivalent to (9). □

Theorem 2. *Interval max-plus matrix equation* (4) *is T3 solvable if and only if interval max-plus matrix equation* (4) *is T10 solvable and*

$$\underline{A} \otimes X^{(uv)} \otimes \underline{C} = \overline{A} \otimes X^{(uv)} \otimes \overline{C} \tag{10}$$

for each $u \in N$, $v \in S$.

Proof. Suppose that $\underline{A} \otimes X^{(uv)} \otimes \underline{C} = \overline{A} \otimes X^{(uv)} \otimes \overline{C}$ for each $u \in N$, $v \in S$ and interval max-plus matrix equation (4) is T10 solvable. According to Lemma 8 we get

$$\underline{A} \otimes X \otimes \underline{C} = \underline{A} \otimes \Big(\bigoplus_{u \in N, v \in S} (x_{uv} - \overline{x}_{uv}) \otimes X^{(uv)} \Big) \otimes \underline{C} = \bigoplus_{u \in N, v \in S} (x_{uv} - \overline{x}_{uv}) \otimes (\underline{A} \otimes X^{(uv)} \otimes \underline{C}) =$$

$$= \bigoplus_{u \in N, v \in S} (x_{uv} - \overline{x}_{uv}) \otimes (\overline{A} \otimes X^{(uv)} \otimes \overline{C}) = \overline{A} \otimes \Big(\bigoplus_{u \in N, v \in S} (x_{uv} - \overline{x}_{uv}) \otimes X^{(uv)} \Big) \otimes \overline{C} = \overline{A} \otimes X \otimes \overline{C}$$

for each $x \in \boldsymbol{X}$. By Lemma 2 interval system (4) is T3 solvable.

The converse implication is easy. □

Lemma 3. Interval max-plus matrix equation (4) is T8 solvable if and only if interval max-plus matrix equation (4) is T10 solvable and

$$A \otimes \underline{X} \otimes C = A \otimes \overline{X} \otimes C \tag{11}$$

for every $A \in \boldsymbol{A}$ and for every $C \in \boldsymbol{C}$.

Proof. The T10 solvability and the monotonicity of $\otimes$ imply that $A \otimes \underline{X} \otimes C \geq \underline{B}$ and $A \otimes \overline{X} \otimes C \leq \overline{B}$ for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$. Let $A \in \boldsymbol{A}$ and $C \in \boldsymbol{C}$ be arbitrary, but fixed. Then the existence of $B \in \boldsymbol{B}$ such that for each $X \in \boldsymbol{X}$ the equality $A \otimes X \otimes C = B$ is satisfied means that the products $A \otimes X \otimes C$ are the same for each $X \in \boldsymbol{X}$, which is equivalent to (11). □

Theorem 3. *Interval max-plus matrix equation* (4) *is T8 solvable if and only if interval max-plus matrix equation* (4) *is T10 solvable and the equality*

$$A^{(ij)} \otimes \underline{X} \otimes C^{(lk)} = A^{(ij)} \otimes \overline{X} \otimes C^{(lk)} \tag{12}$$

holds for each $i \in M$, $j \in N$, $k \in R$, $l \in S$.

Proof. Suppose that equality (12) is satisfied for each $i \in M$, $j \in N$ and interval max-plus matrix equation (4) is T10 solvable. Using Lemma 8 (ii) we have

$$A \otimes \underline{X} \otimes C = \Big(\bigoplus_{i,j} (a_{ij} - \overline{a}_{ij}) \otimes A^{(ij)} \Big) \otimes \underline{X} \otimes \Big(\bigoplus_{l,k} (c_{lk} - \overline{c}_{lk}) \otimes C^{(lk)} \Big) =$$

$$= \Big(\bigoplus_{i,j} (a_{ij} - \overline{a}_{ij}) \otimes A^{(ij)} \otimes \underline{X} \Big) \otimes \Big(\bigoplus_{l,k} (c_{lk} - \overline{c}_{lk}) \otimes C^{(lk)} \Big) =$$

$$= \bigoplus_{i,j,k,l} \Big(((a_{ij} - \overline{a}_{ij}) \otimes (c_{lk} - \overline{c}_{lk})) \otimes A^{(ij)} \otimes \underline{X} \otimes C^{(lk)} \Big) = \bigoplus_{i,j,k,l} \Big(((a_{ij} - \overline{a}_{ij}) \otimes (c_{lk} - \overline{c}_{lk})) \otimes A^{(ij)} \otimes \overline{X} \otimes C^{(lk)} \Big) =$$

$$= \Big(\bigoplus_{i,j} (a_{ij} - \overline{a}_{ij}) \otimes A^{(ij)} \Big) \otimes \overline{X} \otimes \Big(\bigoplus_{l,k} (c_{lk} - \overline{c}_{lk}) \otimes C^{(lk)} \Big) = A \otimes \overline{X} \otimes C$$

Thus, by Lemma 3, interval max-plus matrix equation (4) is T8 solvable. □

Theorem 4. *Interval max-plus matrix equation* (4) *is T9 solvable if and only if interval max-plus matrix equation* (4) *is T10 solvable and*

$$\underline{A} \otimes \underline{X} \otimes \underline{C} = \overline{A} \otimes \overline{X} \otimes \overline{C}. \tag{13}$$

Proof. If interval max-plus matrix equation (4) is T10 solvable and $\underline{A} \otimes \underline{X} \otimes \underline{C} = \overline{A} \otimes \overline{X} \otimes \overline{C} = B$, then $B = \underline{A} \otimes \underline{X} \otimes \underline{C} \leq A \otimes X \otimes C \leq \overline{A} \otimes \overline{X} \otimes \overline{C} = B$ for every $X \in \boldsymbol{X}$. We get $A \otimes X \otimes C = B \in \boldsymbol{B}$ for every $A \in \boldsymbol{A}$, for every $C \in \boldsymbol{C}$ and for every $X \in \boldsymbol{X}$. Thus, there exists the matrix B such that for every $A \in \boldsymbol{A}$, for every $C \in \boldsymbol{C}$ and for every $X \in \boldsymbol{X}$, the equality $A \otimes X \otimes C = B$ is true.

The converse implication is easy to see. □

Theorems 1, 4, 3, and 2 give an effective algorithms for checking T10, T9, T8, and T3 solvability for a given interval max-plus matrix equation. In the next theorem, we suppose that $m = r = s = n$.

Theorem 5. *There are algorithms which decide whether the given interval max-plus matrix equation is*
- *T10 solvable or T9 solvable in $O(n^3)$ steps,*
- *T3 solvable in $O(n^5)$ steps,*
- *T8 solvable in $O(n^7)$ steps.*

Proof. We need $O(n^3)$ arithmetic operations for matrix multiplications. Checking matrix inequalities requires $O(n^2)$ arithmetic operations. To determine wheather the given interval max-plus matrix equation is or not T10 solvable, it is necessary to verify inequalities (6), (7). Hence, the complexity is $O(n^3) + O(n^2) = O(n^3)$.

To check the T3 solvability, we need to verify the equation (10) for n^2 ordered pairs (u, v). So, the complexity is $n^2 \cdot O(n^3) = O(n^5)$.

Similarly, the complexity of checking T8 solvability, using Theorem 3, is $n^4 \cdot O(n^3) = O(n^7)$. □

Example 2. Let

$$\boldsymbol{A} = \begin{pmatrix} [2,2] & [-3,-2] \\ [-1,0] & [1,1] \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} [1,2] & [2,3] \\ [0,3] & [1,3] \end{pmatrix}, \boldsymbol{C} = \begin{pmatrix} [1,1] & [1,1] \\ [0,1] & [1,1] \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} [-1,-1] & [-2,-1] \\ [0,0] & [-2,-1] \end{pmatrix}.$$

Using previous theorems, we determine T10, T9, T8, and T3 solvability of interval matrix equation $A \otimes X \otimes C = B$.

Solution: At first, we use Theorem 1 and Theorem 4. We obtain

$$\underline{A} \otimes \underline{X} \otimes \underline{C} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \geq \underline{B}, \quad \overline{A} \otimes \overline{X} \otimes \overline{C} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \leq \overline{B}, \quad \underline{A} \otimes \underline{X} \otimes \underline{C} = \overline{A} \otimes \overline{X} \otimes \overline{C}.$$

Thus, the given interval matrix equation is T10 solvable and also it is T9 solvable.

Next, we compute

$$X^{(11)} = \begin{pmatrix} -1 & -2 \\ 0 & -2 \end{pmatrix}, \quad X^{(12)} = \begin{pmatrix} -1 & -1 \\ 0 & -2 \end{pmatrix}, \quad X^{(21)} = \begin{pmatrix} -1 & -2 \\ 0 & -2 \end{pmatrix}, \quad X^{(22)} = \begin{pmatrix} -1 & -2 \\ 0 & -1 \end{pmatrix}.$$

We calculate the following products

$$\underline{A} \otimes X^{(11)} \otimes \underline{C} = \overline{A} \otimes X^{(11)} \otimes \overline{C} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = \underline{A} \otimes X^{(12)} \otimes \underline{C} = \overline{A} \otimes X^{(12)} \otimes \overline{C} =$$

$$= \underline{A} \otimes X^{(21)} \otimes \underline{C} = \overline{A} \otimes X^{(21)} \otimes \overline{C} = \underline{A} \otimes X^{(22)} \otimes \underline{C} = \overline{A} \otimes X^{(22)} \otimes \overline{C}.$$

Since the equation (10) is satisfied for all $j \in N$, $l \in S$, we apply the Theorem 2, so, the given interval system is T3 solvable.

At last, using Theorem 3, we show that the given interval system is also T8 solvable.

$$A^{(11)} = \begin{pmatrix} 2 & -3 \\ -1 & 1 \end{pmatrix} \quad A^{(12)} = \begin{pmatrix} 2 & -2 \\ -1 & 1 \end{pmatrix} \quad A^{(21)} = \begin{pmatrix} 2 & -3 \\ 0 & 1 \end{pmatrix} \quad A^{(22)} = \begin{pmatrix} 2 & -3 \\ -1 & 1 \end{pmatrix},$$

and

$$C^{(11)} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad C^{(12)} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad C^{(21)} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad C^{(22)} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Since, for every $i \in M$, $j \in N$, $k \in R$, $l \in S$ the products

$$A^{(ij)} \otimes \underline{X} \otimes C^{(lk)} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \quad \text{and} \quad A^{(ij)} \otimes \overline{X} \otimes C^{(lk)} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

are the same, the given interval system is T8 solvable.

References

[1] Myšková, H.: Interval systems of max-separable linear equations. *Lin. Algebra Appl.* **403** (2005), 263–272.

[2] Myšková, H.: Max-min interval systems of linear equations with bounded solution. *Kybernetika* **48**, no. 2 (2012), 299–308.

[3] Myšková, H.: Interval max-plus matrix equations, *Lin. Algebra Appl.* **492** (2016),111–127.

[4] Myšková, H.: Interval fuzzy matrix equations, *Proceeding of 34th International Conference Mathematical Methods in Economics*, (Kocourek, A., Vavrousek, M. eds.), Technical University of Liberec, 2016, 588–593.

[5] J. Plavka, On the $O(n^3)$ algorithm for checking the strong robustness of interval max-plus matrices, Discrete Applied Mathematics 160 (2012), 640 – 647.

[6] Zimmermann, K.: *Extremální algebra*, Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV, Praha, 1976.

Diversification problem in mean-variance-skewness portfolio models

Renata Dudzińska-Baryła[1], Donata Kopańska-Bródka[2], Ewa Michalska[3]

Abstract. In Markowitz's portfolio selection model only two parameters of distributions of rate of return are taken into account. In practice those distributions are mainly asymmetric, so the models which take into account only the mean and variance lack important features of portfolios. We propose a portfolio selection model which maximises the expected value and skewness and minimises the variance. Parametric analysis of the level of diversification allows us to study the stability of the structure of optimal portfolios in relation to the investor preferences regarding the expected return and skewness. In this paper we determine the optimal portfolios of stocks listed on Warsaw Stock Exchange, which have the same degree of diversification. We show that consideration of skewness in the efficient portfolio analysis changes the degree of diversification of optimal portfolios. The greater the strength of preferences for skewness the lower the degree of portfolio diversification.

Keywords: skewness, diversification, stock portfolio.

JEL Classification: G11
AMS Classification: 91G10

1 Introduction

The classical Markowitz approach to portfolio optimization bases on the mean-variance criterion and the assumption that random rates of return are normally distributed. In investment practice when asymmetric distributions are observed the accuracy of model which bases on only the first two moments of rate of return's distribution is questionable. Several alternative approaches to incorporate the asymmetry of the distribution of rates of return in investment decisions is widely considered, e.g. in [1, 20]. The preferences of a decision maker who chooses from a set of random variants with asymmetric distributions can be expressed by the higher order moments of the probability distributions. Decision makers prefer those variants that correspond to the higher values of odd moments (expected value, skewness) and lower values of even moments (variance, kurtosis).

Most of distributions of rates of return are asymmetric and have fat tails which corresponds to large skewness or kurtosis. Maximization of skewness expresses the basic preferences of a decision maker, which is an increase of the chances of achieving above-average rates of return. Skewness measured by the third central moment has been considered in the models of optimal portfolios since the early 1970s [3, 8, 14, 11, 19]. However, the development of computer-aided numerical tools has led to increasing the number of publications in which authors use higher order moments of rates to analyze optimal portfolios [2, 6, 7, 9, 10, 13, 15, 16, 17].

The mean-variance criterion states that investors should always choose an efficient portfolio. In the case where only the first two moments are considered and the short sale is not allowed the efficient frontier is a curve in two-dimensional space and can be defined by corner portfolios. These portfolios divide the efficient frontier into sets of portfolios of the same structure, which means that the degree of diversification of portfolios in a given subset (measured by the number of elements with positive share) is the same, and this division depends only on the non-negative values of a certain parameter [5]. For the portfolio selection model with the first three moments in the objective function the efficient frontier is a surface in the three-dimensional space for which the definition of corner portfolios does not make sense. However, it remains the issue of identifying such subsets of portfolios from this surface for which the degree of diversification is the same, so the purpose of the article is to analyze the degree of diversification of the efficient portfolios in the three-dimensional space. The problem will be illustrated by an empirical example based on data from the Warsaw Stock Exchange.

[1] University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, dudzinska@ue.katowice.pl.
[2] University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, broda@ue.katowice.pl.
[3] University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, ewa.michalska@ue.katowice.pl.

2 Models for selecting efficient portfolios

Potential components of the portfolios are stocks whose random rates of return R_i ($i = 1, ..., N$) form a vector $\mathbf{R} = [R_1, R_2, ..., R_N]$, assuming that the expected values $\mathrm{E}[R_i] < \infty$. Vector of stock shares in the portfolio is $\mathbf{x} = [x_1, x_2, ..., x_N]$ for $x_1 + x_2 + ... + x_N = 1$ and $x_i \geq 0$ for $i = 1, ..., N$. The rate of return of a portfolio is a random variable with the distribution generated by the random rates of return of portfolio components and equals $R_P = R_1 x_1 + R_2 x_2 + ... + R_N x_N$.

The expected value of the portfolio random rate of return ($\mathrm{E}(R_P)$) is the first order moment, while the variance of the portfolio (V_P) is the second central moment of the portfolio random rate of return. As a measure of skewness of the portfolio random rate of return the third central moment is assumed. The skewness of portfolio can be written as

$$\mathrm{S}_P = \mathrm{E}[\mathbf{R_p} - \mathrm{E}(R_p)]^3 = \mathbf{x} \cdot \mathbf{M}_3 \cdot (\mathbf{x}^{\mathrm{T}} \otimes \mathbf{x}^{\mathrm{T}}) \tag{1}$$

where the symbol $\otimes$ denotes the Kroneckera product. Elements of the $(N \times N^2)$ matrix $\mathbf{M}_3 = \mathrm{E}[(\mathbf{R} - \mathrm{E}(\mathbf{R})) \cdot (\mathbf{R} - \mathrm{E}(\mathbf{R}))^{\mathrm{T}} \otimes (\mathbf{R} - \mathrm{E}(\mathbf{R}))^{\mathrm{T}})]$ are the third central moments and the co-skewnesses of the random rates of return of stocks. The measure of the skewness of the random rate of return of the portfolio can also be presented as the following sum

$$\mathrm{S}_P = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} s_{ijk} x_i x_j x_k \tag{2}$$

where $s_{ijk} = \mathrm{E}[(R_i - \mathrm{E}(R_i))(R_j - \mathrm{E}(R_j))(R_k - \mathrm{E}(R_k))]$, for $i, j, k = 1, ..., N$.

The two objective problem of maximizing the expected value and minimizing variance is reduced to a one objective model of maximizing the linear utility function with respect to the first two moments of the distribution of portfolio rate of return in the form

$$\mathrm{u}(\mathrm{E}, \mathrm{V}) = \lambda \mathrm{E}(R_P) - \mathrm{V}_P \tag{3}$$

where parameter λ is non-negative. The function (3) is increasing with respect to the expected value and decreasing with respect to the variance. The set of efficient portfolios can be determined by solving a quadratic optimization model:

$$\begin{aligned} &\lambda \mathrm{E}(R_P) - \mathrm{V}_P \rightarrow \max \\ &\sum_{i=1}^{N} x_i = 1 \\ &x_i \geq 0, \qquad i = 1,...,N \end{aligned} \tag{4}$$

by changing the value of parameter λ in interval $\langle 0; +\infty)$. However, to determine the entire efficient frontier it is enough to determine the set of corner portfolios. The necessary condition for two efficient portfolios to be adjacent corner portfolios is that their compositions vary by exactly one share included or excluded. The methodology of determining the corner portfolios has been presented in details in [5, 12, 18].

Changing the value of parameter λ the solutions of problem (4) are efficient portfolios and their parameters $\mathrm{E}(R_P)$ and V_P determine the line of the efficient frontier. The ranges of parameter λ for which efficient portfolios have the same structure can be specified. If an investor is interested only in portfolios that contain K shares then it is sufficient to choose the appropriate segment of the efficient frontier between two adjacent corner portfolios, one of which is made up of K shares and the other of $K-1$ shares. If, however, the investor's preference is expressed by assuming values of the first or second moment then it is possible to determine the degree of diversification of the portfolio based on the structure of the corner portfolios. The graphic illustration of such situation is presented in Fig. 1. The corner portfolios are marked with dots. If the investor prefers portfolio with expected return μ then his efficient portfolio (Fig. 1a) contains as many shares as portfolio P_m or P_{m+1}. Similarly, the investor can determine the degree of portfolio diversification when he defines his risk preferences at level σ (Fig. 1b).

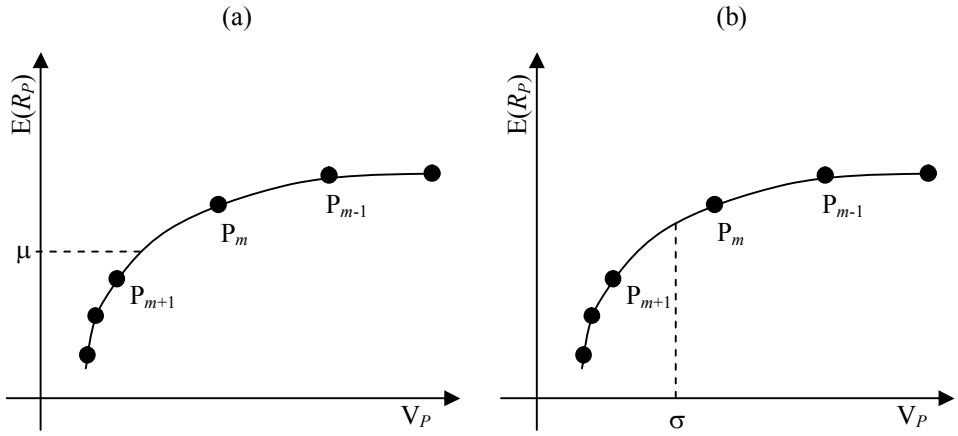


Figure 1 Efficient portfolios that correspond to the investor's preference for the expected value or variance

The problem of selecting efficient portfolio which both maximizes the expected return and skewness as well as minimizes the variance can be stated as:

$$\begin{aligned} &\lambda \mathrm{E}(R_P) - \mathrm{V}_P + \gamma \mathrm{S}_P \to \max \\ &\sum_{i=1}^{N} x_i = 1 \\ &x_i \geq 0 \qquad \text{dla} \quad i = 1, \ldots, N \end{aligned} \tag{5}$$

where $\lambda \geq 0$ and $\gamma \geq 0$.

If $\lambda = 0$ the solutions of the model (5) are efficient portfolios in the risk-skewness plane. For $\gamma = 0$ the Markowitz's model (4) is obtained. In turn, for $\lambda = 0$ and $\gamma = 0$ a global minimum risk portfolio is the solution of the model (5).

For $\lambda > 0$ and $\gamma > 0$ the objective in the model (5) takes into account all three moments of the portfolio random rate of return and the efficient frontier is the surface in the three-dimensional space. Determining the surface is subject to great numerical difficulties because it requires solving problems that are not convex and that surface cannot be characterized by corner portfolios. The authors of [4] have proposed some solution to this problem and the geometrical interpretation of efficient portfolios in the three-dimensional space.

3 Diversification of optimal portfolios with mean-variance-skewness criterion

In this study on the optimal structure of portfolios based on the model (5) our data set contains daily returns from the first quarter of 2016 for 13 shares listed on the Warsaw Stock Exchange. These shares are components of the WIG20 index. Calculations were made in the SAS software using one of the NLPC solver algorithm - Newton-type method with linear search and self-prepared programs.

Values of parameters λ and γ in model (5) were from intervals $\langle 0; 140 \rangle$ and $\langle 0; 85 \rangle$ with precision 0.5. For values beyond these upper limits only one-share portfolios were obtained.

When the investor takes into account two characteristics of portfolios, i.e. expected return and risk or skewness and risk, the efficient frontier is the curve and it is relatively easy to relate the degree of diversification of efficient portfolios with those parameters.

First, we analyzed the degree of diversification of the efficient portfolios in the risk-return plane. Table 1 presents the relationship between the number of shares constituting the efficient portfolios and intervals of their expected returns and variances. For efficient portfolios obtained for the variance-skewness criterion (when $\lambda = 0$) the relationship between the degree of diversification and the parameters of distribution are presented in table 2.

Degree of diversification K	Expected return $E(R_P)$	Variance V_P	Degree of diversification K	Expected return $E(R_P)$	Variance V_P
2	0.4185-0.4515	5.1891-8.3144	8	0,2537-0,2559	1,1891-1,2061
3	0.4132-0.4184	4.9308-5.1890	9	0.2402-0.2536	1.1003-1.1890
4	0.3679-0.4131	3.2009-4.9307	10	0.2165-0.2401	0.9866-1.1002
5	0.3648-0.3678	3.1122-3.2008	11	0.1699-0.2164	0.8806-0.9865
6	0.3565-0.3647	2.8880-3.1121	12	0.1512-0.1698	0.8723-0.8805
7	0.2560-0.3564	1.2062-2.8879	13	N/A	N/A

Table 1 Degree of diversification, expected return and variance for efficient portfolios based on (5) for $\gamma = 0$

Degree of diversification K	Skewness S_P	Variance V_P	Degree of diversification K	Skewness S_P	Variance V_P
2	5.6564-5.7197	3.7401-3.9642	8	0.9062-1.2563	0.9543-1.0788
3	4.6860-5.6563	2.8072-3.7400	9	0.7656-0.9061	0.9178-0.9542
4	3.6369-4.6859	2.1445-2.8071	10	0.6021-0.7655	0.8876-0.9177
5	3.5421-3.6368	2.0964-2,1444	11	0.5315-0,6020	0.8795-0.8875
6	2.7740-3.5420	1.7317-2.0963	12	0.3813-0.5314	0.8723-0.8794
7	1.2564-2.7739	1.0789-1.7316	13	N/A	N/A

Table 2 Degree of diversification, skewness and variance for efficient portfolios based on (5) for $\lambda = 0$

To analyze the diversification of optimal portfolios based on the model (5) the map of the degree of diversification of portfolios can be used. Figure 2 shows an example of a map of the degree of diversification of efficient portfolios for a selected intervals of parameters $\lambda \in \langle 0;30 \rangle$ and $\gamma \in \langle 0;1.5 \rangle$. Single-color areas correspond to portfolios of the same structure, for example the black area corresponds to the most diversified portfolios (at least 10 shares). For the selected degree of diversification of the portfolio, the investor can read the range of corresponding values of the parameters λ and γ.

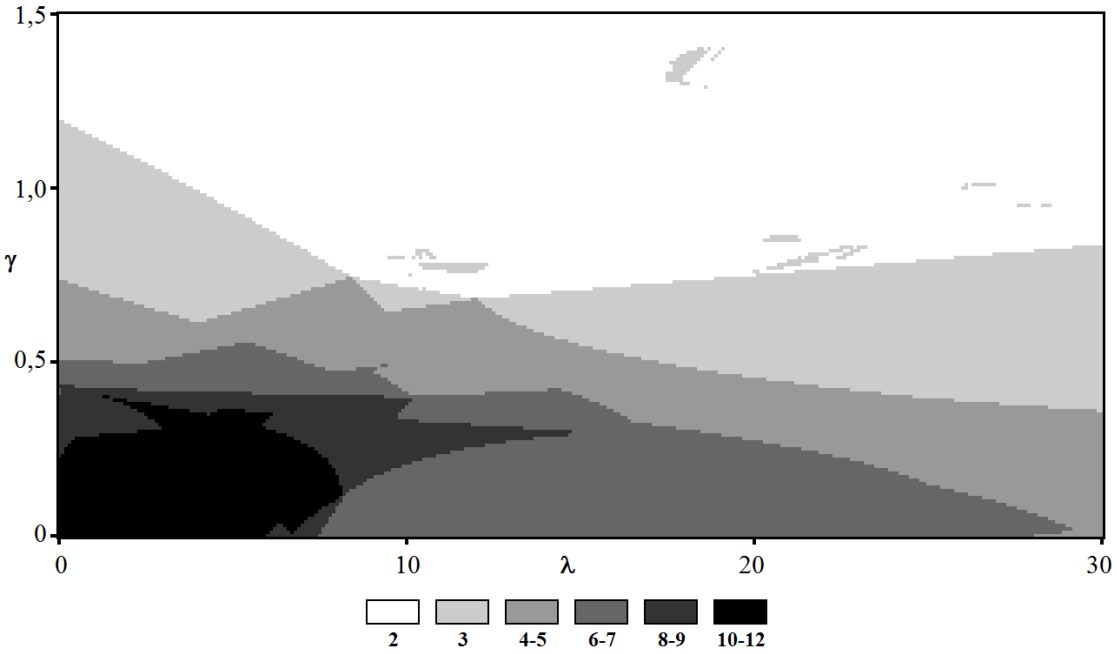


Figure 2 Map of the degree of the diversification of optimal portfolios for $\lambda \in \langle 0;30 \rangle$ and $\gamma \in \langle 0;1.5 \rangle$

Based on the map in Figure 2, it is possible to analyze the relationship between the strength of the preference with respect to the rate of return or the skewness and the degree of portfolio diversification. The strength of the

investor's preference for a portfolio's rate of return is expressed by the value of the parameter λ, while the value of the parameter γ expresses the strength of the preference for the skewness. For a fixed value of γ e.g. 0.4, with the increase of the strength of preference for the portfolio's rate of return (the increase in the parameter λ), the degree of diversification of the portfolio changes by one share added or excluded (Fig. 3a), similarly as for corner portfolios. Generally, it can be said that more diversified portfolios correspond to smaller values of λ. Similar conclusions can be drawn for a fixed value of parameter λ e.g. $\lambda = 6$ (Fig. 3b). The low strength of preferences for skewness corresponds to the more diversified portfolios (9 to 11 shares), and with the increase in the strength of preference for the third moment, the degree of portfolio diversification is rapidly decreasing.

(a) for $\gamma = 0{,}4$ (b) for $\lambda = 6$

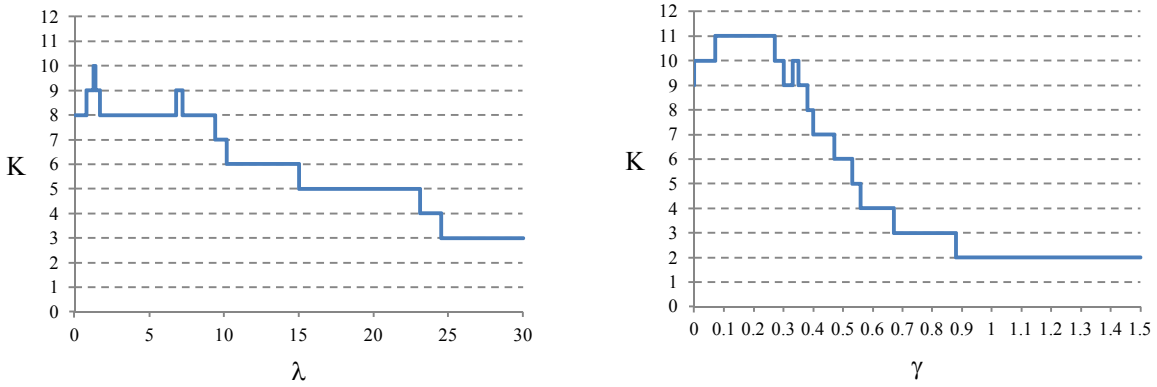


Figure 3 The degree of the diversification of optimal portfolios for mean-variance-skewness criterion

For any assumed levels of strength of preference for the expected rate of return of the portfolio or the skewness (expressed by the values of parameters λ and γ), it is possible to graph the degree of diversification of effective portfolios with respect to the strength of preference for skewness or rate of return.

4 Conclusions

In this study our main objective is to show the relationship between the degree of diversification and the parameters of the efficient portfolios. The higher order moments of the probability distribution are necessary in an adequate analysis of the portfolios selection. The use of only the expected value and variance is justified if those characteristics are sufficient to determine the probability distribution function. This applies only to normal and uniform distribution.

The need to take into account the skewness in choosing the optimal portfolio has its empirical and theoretical justification. In most cases, distributions of rates of return are asymmetric. Maximization of skewness expresses the basic preferences of a decision maker, which is an increase of the chances of achieving above-average rates of return. However, determining an efficient frontier for an optimization model in which the objective function depends on three moments of probability distribution is difficult and rarely performed due to the lack of a suitable algorithm.

One of the basic steps in the portfolio construction is to determine the structure and thus to determine the number of shares which form a portfolio. Linking the degree of diversification of efficient portfolios to the corresponding values of the model parameters (5) allows the investor to identify the optimal portfolio that corresponds to his or her preference for the portfolio structure. Taking into account in the portfolio analysis the third central moments of the distributions influences the degree of diversification. The greater the strength of preference for skewness, the more diversified the portfolios are. Also, with the increase in the value of the parameters λ and γ the number of components in the portfolios changes by one share.

References

[1] Arditti, F. D.: Skewness and Investor's Decisions: A Reply. *The Journal of Financial and Quantitative Analysis* **10** (1975), 173–176.

[2] Athayde, G., Flores, R.: Finding a Maximum Skewness Portfolio – A General Solution to Three-Moments Portfolio Choice. *Journal of Economic Dynamics and Control* **28** (2004), 1335–1352.

[3] Barone-Adesi, G.: Arbitrage Equilibrium with Skewed Asset Returns. *Journal of Financial and Quantitative Analysis* **20** (1985), 299–313.
[4] Briec, W., Kerstens, K., Van de Woestyne, I.: Portfolio selection with skewness: A comparison of methods and a generalized one fund result. *European Journal of Operational Research* **230** (2013), 412–421.
[5] Dudzińska, R.: Wykorzystanie portfeli rogowych do wyznaczania linii efektywnej. In: *Modelowanie preferencji a ryzyko '02* (T. Trzaskalik, ed.). Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice, 2002, 93-106.
[6] Dudzińska-Baryła, R., Kopańska-Bródka, D., Michalska, E. Analiza portfeli narożnych z uwzględnieniem skośności. *Zeszyty Naukowe Uniwersytetu Szczecińskiego nr* **862**. *Finanse, Rynki Finansowe, Ubezpieczenia nr* **75** (2015), 123–133.
[7] Guidolin, M., Timmermann, A.: International Asset Allocation under Regime Switching, Skew, and Kurtosis Preferences. *Review of Financial Studies* **21** (2008), 889–935.
[8] Kane, A.: Skewness Preference and Portfolio Choice. *Journal of Financial and Quantitative Analysis* **17** (1982), 15–25.
[9] Kemalbay, G., Özkut, C. M., Franko, C.: Portfolio Selection with Higher Moments: A Polynomial Goal Programming Approach to ISE-30 Index. *Ekonometri ve İstatistik Sayı* **13** (2011), 41–61.
[10] Kim, W. Ch., Fabozzi, F. J., Cheridito, P., Fox, Ch.: Controlling Portfolio Skewness and Kurtosis Without Directly Optimizing Third and Fourth Moments. *Economic Letters* **122** (2014), 154–158.
[11] Konno, H., Shirakawa, H., Yamazaki, H.: A Mean-absolute Deviation-skewness Portfolio Optimization Model. *Annals of Operations Research* **45** (1993), 205–220.
[12] Kopańska-Bródka, D. (ed.): *Wybrane problemy ilościowej analizy portfeli akcji*. Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice, 2004.
[13] Kopańska-Bródka, D.: Optymalny portfel inwestycyjny z kryterium maksymalnej skośności. *Studia Ekonomiczne. Zeszyty Naukowe Wydziałowe Uniwersytetu Ekonomicznego w Katowicach* **208** (2014), 46–58.
[14] Lai, T.: Portfolio Selection with Skewness: A Multi-Objective Approach. *Review of Quantitative Finance and Accounting* **1** (1991), 293–305.
[15] Li, X., Qin, Z., Kar, S.: Mean-variance-skewness Model for Portfolio Selection with Fuzzy Returns. *European Journal of Operational Research* **202** (2010), 239–247.
[16] Malevergne, Y., Sornette, D.: Higher-Moment Portfolio Theory. *Journal of Portfolio Management* **31** (2005), 49–55.
[17] Proelss, J., Schweizer, D.: Polynomial Goal Programming and the Implicit Higher Moment Preferences of US Institutional Investors in Hedge Funds. *Financial Markets and Portfolio Management* **28** (2014), 1–28.
[18] Sharpe, W. F.: *Macro-Investment Analysis: Optimization. The Critical Line Method*. (2000) (http://web.stanford.edu/~wfsharpe/mia/opt/mia_opt3.htm)
[19] Simonson, D.: The Speculative Behavior of Mutual Funds. *Journal of Finance* **27** (1972), 381-391.
[20] Xiong, J. X., Idzorek, T. M.: The Impact of Skewness and Fat Tails on the Asset Allocation Decision. *Financial Analysts Journal* **67** (2011), 23–35.

Quantification of latent variables based on relative interaction between players

Marek Dvořák [1], Petr Fiala[2]

Abstract. In this paper, we propose a method for estimating a relative player ranking based on a series of finished games played in a bigger tournament. The players 'score' in a game is an unobservable latent variable and the only way to measure it is indirectly through a result of a match played between two players. In these matches a definite winner can be determined, thus we do not consider games which could end up with a tie. We have also developed a model to measure tournament transitivity and player consistency. Since the player ranking is a one dimensional value, their ordering has to be transitive, but the results of the tournament in general may not conform to this notion. This would result in estimations with very high variance. Our measures take both the transitivity and player estimation variance into account. We have also applied those models to several real world tournaments and their results to test our approach on real data.

Keywords: Player ranking, latent variable, transitivity, consistency, probability

JEL classification: C44
AMS classification: 90C15

1 Introduction

Player raking systems are a branch of multi criteria decision making system which aim to accurately describe players based on results of series of games already played between those players. The goal is usually to estimate some latent variable that cannot be directly observable, or such observation is very difficult. The only available metric is usually a result of one match between two players, but their individual score in this game is usually unknown. What is more, such score is also usually a result of interaction between two such competing players and this is highly relative and directly unusable in absolute player quality quantification. A player is described using several variables which define a probability distribution because his individual performance is probabilistic in nature.

Many tournaments are basically setting themselves to determine the relative comparison between such latent variables which often results in some sort of rudimentary sorting. This is usually determined by round robin or elimination tournaments from which a winner is determined with runner-ups alongside. This is often a rudimentary ordinal sorting lacking any cardinal information. With elimination tournaments, the lower places may lack even such distinction.

Many algorithms were devised to estimate such latent player variables, most notably ELO [2] or Glicko [3] family ranking algorithms. These types of algorithm however try to estimate player rankings online, over long period of time and and take the long-time shift of latent variables into consideration. This paper begins by proposing an algorithm to determine static latent variables holistically by looking at all the available data at once. Then there is also a demonstration of this algorithm to show it working on a real world tournament that was held.

2 Latent variables

A player i is associated with random normal distribuition:

$$X_i, \ldots, X_n \sim N(\mu_i, \sigma_i^2) \tag{1}$$

A player's individual plays would be samples from this particular distribution x_i^k if they were directly observable. A mean value μ_i is indication of a player's skill, whereas variability σ_i^2 is indication of a player's consistency.

You can easily imagine the whole issue when you think of the player like some sort of athlete sprinter. His average time in a 100m sprint will most likely oscillate around this athlete's mean μ, but no single run will time exactly the same, there will be some variance σ^2 describing his consistency.

[1]University of Economics, Prague, department of econometrics, W. Churchill Sq. 1938/4, Prague, marek.dvorak@vse.cz
[2]University of Economics, Prague, department of econometrics, W. Churchill Sq. 1938/4, Prague, pfiala@vse.cz

However in our case, the player's plays (runs) are not easily measurable and the estimation of variables describing his probability distribution is not trivial.

The variables μ_i and σ_i^2 are latent and not directly observable. We only observe a result of a single match between two players - a nonequivalence operator between samples x_i^k and x_j^k (samples from X_i and X_j), $x_i^k < x_j^k$ or $x_i^k > x_j^k$. We can also observe series of matches between several players, calling it a tournament.

Suppose we have a set N of n players described by quantities:

$$X_1 \sim N(\mu_1, \sigma_1^2),\ \ldots,\ X_n \sim N(\mu_n, \sigma_n^2) \tag{2}$$

We do not directly observe these quantities.

Also suppose these n players participated in a tournament, where m matches were played in total.

We can directly observe a result τ_{ijk} between player i and player j in a kth game of the tournament.

$$\tau_{ijk} = \mathrm{I}\left[x_i^k > x_j^k\right] \tag{3}$$

$\tau_{ijk} = 1$ if ith player won over the jth player in a kth game of the tournament and $\tau_{ijk} = 0$ otherwise. I do not consider the possibility of ties here.

We can also define a tournament matrix $T = (t_{ij}) \in \mathbb{R}^{n\times n}$. Its meaning is „How many times has player i won over a player j?“ and it is a simple sum over all the games in a tournament:

$$t_{ij} = \sum_{k=1}^{\zeta_{ij}} \tau_{ijk} \quad \forall i, j \in N \tag{4}$$

where ζ_{ij} represents how many times has those players played together.

I also make an assumption that individual player's μ_i and σ_i^2 does not change during the tournament.

3 Observations

If I have two players i and j with their respective means $\mu_i; \mu_j$ and variances $\sigma_i^2; \sigma_j^2$, I can express the probability that player i wins over a player j:

$$p_{ij} := \Pr\left[X_i > X_j\right] = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{\sigma_i^2 + \sigma_i^2}}\right) \tag{5}$$

where Φ is a probability density function of a normal distribution.

If t_{ij} means how many times a player i won over a player j with a probability p_{ij}, I can say that t_{ij} is sampled from a binomial distribution:

$$t_{ij} \sim \mathrm{Bi}\left(p_{ij}, \zeta_{ij}\right) \tag{6}$$

4 Estimating player's parameters

Given the tournament matrix t_{ij}, one could try to estimate the latent variables $\mu_1, \ldots, \mu_n, \sigma_1, \ldots, \sigma_n$ describing the players and then make an inference about their rank or other sort of result or ordering together with inference about their consistence.

We can estimate the probability $\widehat{p}_{ij}$ as:

$$\widehat{p}_{ij} = \frac{t_{ij}}{t_{ij} + t_{ji}} \tag{7}$$

Thus the whole problem reduces to solving a system of equations:

$$\frac{\mu_i - \mu_j}{\sqrt{\sigma_i^2 + \sigma_i^2}} = \Phi^{-1}(\widehat{p}_{ij}) \quad \forall i \neq j \in N \tag{8}$$

Finally, we can estimate those parameters using non-linear mathematical programing techniques and minimize following utility function:

$$z = \sum_{i=i}^{n} \sum_{j=1}^{n} \left(\Phi^{-1}(\widehat{p}_{ij}) - \frac{\mu_i - \mu_j}{\sqrt{\sigma_i^2 + \sigma_i^2}} \right) \rightarrow min \tag{9}$$

Proposition 1. *The family of parameters $\mu_1, \ldots, \mu_n, \sigma_1, \ldots, \sigma_n$ is not identified.*

Proof. If we chose two constants a and b, we can transform μ_{ij} and σ_{ij} into μ'_{ij} and σ'_{ij} like so:

$$\mu'_{ij} = a + b\mu_{ij} \tag{10}$$
$$\sigma'_{ij} = b\mu_{ij} \tag{11}$$

We can also define p'_{ij} like:

$$p'_{ij} := \Phi \left(\frac{\mu'_i - \mu'_j}{\sqrt{\sigma'^2_i + \sigma'^2_i}} \right) \tag{12}$$

and we can put that equation in:

$$p'_{ij} = \Phi \left(\frac{(a + b\mu_i) - (a + b\mu_j)}{\sqrt{(b\sigma_i)^2 + (b\sigma_i)^2}} \right) \tag{13}$$

and if we reduce the above we get $p_{ij} = p'_{ij}$. This means that we can move and scale off of μ_i's along the real axis with parameters a and b and the probability p_{ij} remains the same given that we also scale σ_i accordingly.

Thus the system of equations (8) has effectively $2n-2$ of variables we can estimate from n^2-n equations. □

4.1 Transitivity

Since this model considers the latent variables to be one-dimensional, there is a necessary condition on transitivity as seen in 14.

$$\forall \mu_a, \mu_b, \mu_c \in X : (\mu_a > \mu_b \wedge \mu_a > \mu_c) \Rightarrow \mu_b > \mu_c \tag{14}$$

The tournament matrix however cannot be bound by this transitivity requirement, because result like in a table 1 is possible.

In such case, it is considered that any such intransitivity is basically a result of random error which is present thanks to the probabilistic nature of the observations. This intransitivity can be quantified very simply using the value z of the utility function from equation 9 because if the value $z = 0$ then all equations 8 relating the observed probabilities and latent probabilities were perfectly satisfied. Any observed intransitivity results in a unability of utility function to reach zero.

	A	B	C
A		1	
B			1
C	1		

Table 1 Intransitive tournament

5 Real world example

The whole model can be demonstrated working on a real world tournament. The choice was a 2017 eSports Counter-Strike: Global Offensive tournament – ELEAGUE Major 2017 held from 22nd January, 2017 to 29th January, 2017 at Krakow [1]. There, sixteen teams participated and played 52 games in total.

The tournament matrix T (as shown in table 2) was handcrafted from the tournamet results.

	NV	VP	Fn	G	SK	As	FZ	N	En	GOD	Li	G2	mo	Op	Fl	H
Natus Vincere					1	1			1				1			
Virtus.pro				1	2	1		2				1		1		
Fnatic				2				1	1				1			
Gambit Esports			1				1	1		1						
SK Gaming						1	3									1
Astralis	2	2	2								1	1		1		
FaZe Clan					1				1		1				1	
North		1		1								1				1
Team EnVyUs										1	1					
GODSENT						1								1		
Team Liquid													1		1	
G2 Esports			1													
mousesports																1
OpTic Gaming															1	
FlipSid3 Tactics																
HellRaisers																

Table 2 Tournament matrix

The probabilities $\widehat{p}_{ij}$ were then estimated according to the equation 7. Those were then used to estimate parameters $\mu_1, \ldots, \mu_{16}, \sigma_1, \ldots, \sigma_{16}$ using a BOBYQA [5] algorithm implemented in the R language as part of *nloptr* package [4]. The algorithm output a set of variables that shown in the table 3, which can be interpreted as team ranking and consistency. The utility function reached local minima at $z = 264.13$.

Those results can be visualized as a log-normal distribution with the mean value as $\log_e \mu_i$ and standard deviation of σ_i as shown in figure 1.

This can be compared to the actual tournament results, which were: Astralis winning the tournament and Virtus.pro running second, Fnatic and SK Gaming being eliminated at semi finals and Natus Vincere, Gambit Esports, North and FaZe Clan being eliminated at quarter finals. The rest of the teams were eliminated at group stage. It can be seen that the official results match pretty closely with the estimation given by this method, considering fundamentally different approach from the official tournament format.

6 Conclusion

This paper introduced a model to estimate latent variables describing the relative player ability as a probability distribution. The parameters of such distribution were result of a non-linear optimization problem which minimized the disparity between observed win-rate and latent theoretical win rate between pairs of players. The role of transitivity was also briefly mentioned and discussed.

It was shown that this model accurately behaves accurately on a real world data. This data was an eSports tournament held at January, 2017. The differences between modeled ranks and official tournament results were minimal and attributed to different methods of ordinal estimation.

	μ_i	σ_i
Virtus.pro	201.60	1.68
Astralis	181.51	1.79
Natus Vincere	171.17	1.87
FaZe Clan	151.09	1.88
Fnatic	143.23	1.90
Gambit Esports	136.17	1.92
North	135.44	1.88
SK Gaming	122.01	1.92
GODSENT	110.45	2.05
Team EnVyUs	74.80	1.90
Team Liquid	73.94	1.83
G2 Esports	24.42	1.81
OpTic Gaming	22.41	1.76
mousesports	19.98	1.85
FlipSid3 Tactics	10.00	1.89
HellRaisers	10.00	1.86

Table 3 Rank estimation

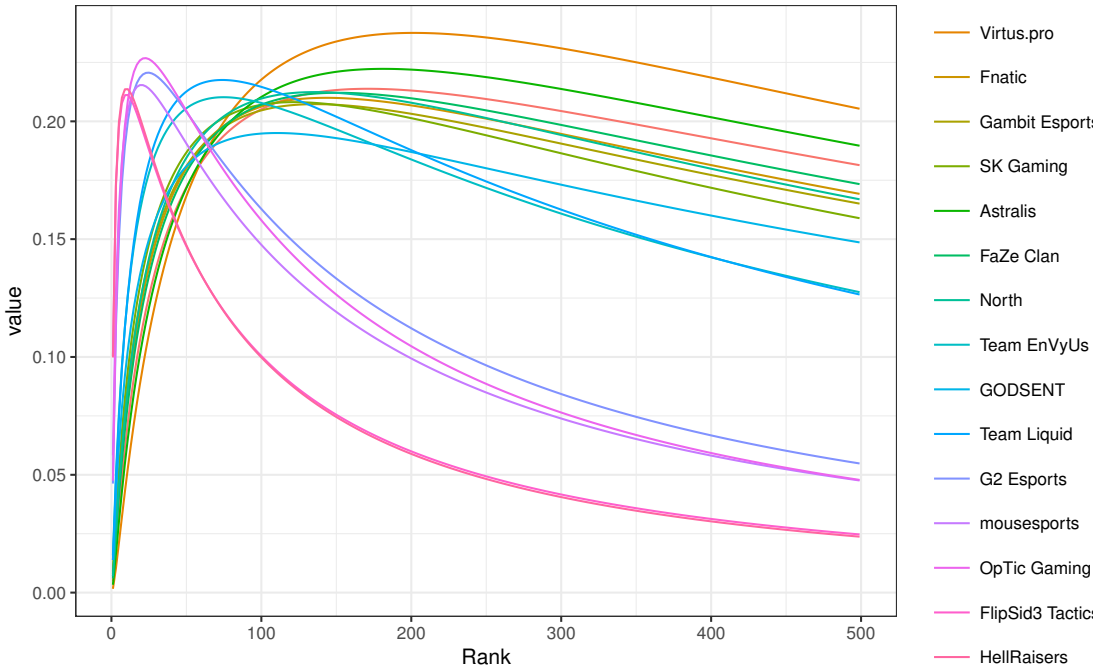


Figure 1 Player distribution visualization

Further improvements could be made by generalizing the model, such as the introduction of tie matches between players, ability to estimate the ranks of individual players based on randomized teams or the notion of latent variable shift in long term. On the other hand, some specialization could also be considered, most notably the correlation between rank and variance of player.

Acknowledgements

Supported by the grant No. IGA F4/57/2017 of the Internal Grant Agency, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] *ELEAGUE Major 2017*, liquidpedia (2017), retrieved from wiki.teamliquid.net/counterstrike/ ELEAGUE/ 2017/ Major

[2] Elo, Arpad E.: *The rating of chessplayers, past and present.* Arco Pub., 1978.

[3] Glickman, Mark E.: *Example of the Glicko-2 system.* Boston University (2012).

[4] Johnson, S. G.: *The NLopt nonlinear-optimization package*, http://ab-initio.mit.edu/nlopt

[5] Powell, M. JD.: The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06*, University of Cambridge, Cambridge (2009).

Goodness of Fit Test for Truncated Distributions, the Empirical Study

Krzysztof Echaust, Agnieszka Lach[1]

Abstract. Since an assumption of normality was rejected on financial markets, many heavy-tailed distributions were pro-posed in literature. Many researchers showed, that semi-heavy tailed distributions are the most suitable to de-scribe stocks or indices behaviour. However, modelling distributions of returns is of less importance around median than it is in the tails, where extreme events appear. Studies of tail thickness confirmed, that they are embraced between Gaussian and alpha-stable models. Our research also concentrates on the tails. We tried to approximate them with several distributions, covering the whole range of possibilities, from thin to fat. Addi-tional value comes from applied goodness of fit tests. When data is truncated, which is our case, testing goodness of fit using standard statistical tests is inappropriate. Instead, we suggest employing modified standard tests like Anderson-Darling, Kolmogorov-Smirnov, Kuiper, Cramér-von Mises ones and tests designed specifically to measure the fit in tails. We conducted research based on 19 assets like indexes, currencies, stocks and future contracts, putting special emphasis on the Polish stock market.

Keywords: Goodness of fit tests, truncated distributions, fat tails.

JEL Classification: C12, C24, C46
AMS Classification: 62G10, 62N01

1 Introduction

The appropriate choice of returns distribution has been a subject of many studies. In the early 1960s Mandelbrot [15] and Fama [9] rejected Gaussian distribution and proposed to use a symmetric Lévy stable distribution as a model for returns behaviour. The empirical evidence concerning Lévy distribution was continued by Fielitz [10], Rachev and Mittnik [19], Rachev [20], Bradley and Taqqu [1], Weron [22], Kabasinskas et al.[13] and many others. In the late 1970s Barndorff-Nielsen [2] proposed Generalized Hyperbolic Distribution (GHD) as a new class of distribution in financial applications. The distribution is widely used in finance, risk management, forecasting and statistical inference. It can be used for unconditional modelling as well as conditional [12]. The empirical evidence on usage of GHD in finance one can find in works of Eberlein i Keller [6], Eberlein et al.[7], Raible [21], Küchler et al.[14]. All the considerations focus on a precise estimation of the distribution tails. They play a key role while downside or upside risk measures, like Value at Risk or Expected Shortfall, are estimated.

Our research also concentrates on the tails of returns distribution. However, instead of modelling the whole returns distribution, we tried to approximate only tails and concentrate on extreme returns. The empirical study takes into account distributions, covering a whole range of possibilities − from thin to fat. An additional value comes from applied goodness of fit tests − when data is truncated, which is in our case, testing goodness of fit using standard statistical tests is inappropriate. Instead, we suggest employing the modified standard tests like Anderson-Darling, Kolmogorov-Smirnov, Kuiper, Cramér-von Mises ones and the tests designed specifically to measure fit in tails, as was proposed in [4]. We conducted the research based on several assets, like indexes, currencies, stocks and future contracts, putting special emphasis on the Polish stock market. The aim of our study is to evaluate the goodness of fit of the theoretical distributions to the empirical ones taking advantage of variety of asset classes. Note that, in our work we concentrate only on the tails of distributions.

The paper is organised as follows: the second section presents some theoretical results about the truncated distribution. It describes the lognormal, Weibull and Generalized Pareto distributions. The third section deals with the Goodness-of-Fit Tests for Truncated Distribution, the fourth section presents the empirical results, and the fifth section is the conclusion section.

[1] Poznan University of Economics and Business, Operations Research Department, al. Niepodległości 10, Poznań: k.echaust@ue.poznan.pl, agalach@interia.pl.

2. Distributions for truncated data

Lognormal Distribution

A random variable X has a lognormal distribution, if the logarithm of the X has a Gaussian distribution. The complementary cumulative distribution of lognormal distribution is defined as follows:

$$\Pr[X \geq x] = \int_x^{\infty} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln z - \mu}{\sigma}\right)^2} dz$$

where: $\mu \in R, \sigma > 0.$ The lognormal distribution is a basic distribution used in option pricing models [3, 17]. The distribution is right-handed skewed and its kurtosis takes values from 3 to the potentially large values depending on σ [5]. If σ is sufficiently large, the logarithm of the density function appears almost linear for a large range of values. The same idea holds for cumulative distribution function. The lognormal distribution is then extremely similar in shape to the power law distributions. The log-log plot of the complementary cumulative distribution function or the density function looks like nearly a straight line for a large portion of the body of the distribution [18].

Weibull Distribution

The Weibull distribution is the distribution with CDF defined as

$$F(x) = 1 - e^{-\frac{1}{2}\left(\frac{x}{\beta}\right)^{\eta}},$$

where $x > 0$ and $\beta, \eta > 0.$

The distribution is right skewed and its density tends exponentially fast to zero, thus Weibull distribution is a thin tailed distribution.

Generalized Pareto Distribution

The GPD is a class of distributions derived from the Extreme Value Theory (EVT). On the basis of the EVT and Balkema-de-Haan theorem, the GPD is a limiting distribution for return exceedances over a high threshold. The broad description of the EVT we can find in [8].The cumulative distribution function of the Generalized Pareto Distribution of the form:

$$F(y) = 1 - \left(1 + \xi \frac{y}{\beta}\right)^{-\frac{1}{\xi}},$$

where: $1 + \xi \frac{y}{\beta} > 0,$ and $\beta > 0$, $y \geq 0$ for $\xi \geq 0$ and $0 \leq y \leq -\beta/\xi$ for $\xi < 0$.The most important parameter is the shape parameter ξ. Depending on the sign of the shape parameter $Y = \frac{x-\mu}{\sigma}$we can obtain a fat-tailed distribution when the shape parameter is positive (Fréchet domain of attraction), a thin-tailed distribution when the shape parameter is equal to zero (the Gumbel domain of attraction) and distribution with a finite endpoint when the shape parameter is equal to zero (the Weibull domain of attraction).

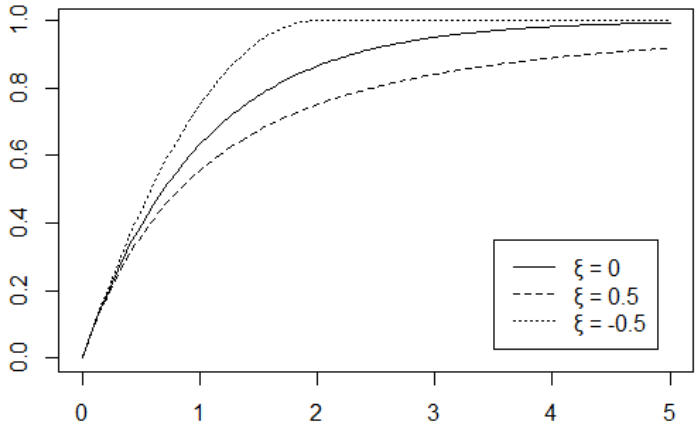


Figure 1. CDF of the Generalized Pareto Distribution for various shape parameter values and $\beta = 1$

3. Goodness-of-Fit Tests for Truncated Distribution

Testing the goodness of fit of truncated data is usually done by appropriate adjustments made to the standard GOF tests or to the sample itself. A brief overview of other possibilities might be found in [11]. This research concentrates on the specific modifications to the standard GOF tests and the two new statistics proposed in [4] to test the goodness of fit for the left-truncated samples.

Let's assume that we have a sample of *n* observations, each exceeding a certain threshold *H*. The standard empirical distribution function determined on such truncated sample will be denoted by $F_n(x)$. We also need $\hat{F}_\theta(x)$, the fitted distribution function for the complete sample. The vector θ of the parameters is unknown and will be estimated with the maximum likelihood method, conditionally on the sample we possess. The modified fitted distribution function for the truncated sample is then as follows:

$$\hat{F}^*(x) = \begin{cases} \dfrac{\hat{F}_\theta(x) - \hat{F}_\theta(H)}{1 - \hat{F}_\theta(H)} & x \geq H \\ 0 & x < H. \end{cases},$$

We will verify the composite hypothesis, whether the sample comes from the population characterized by the specified truncated continuous distribution in the following form:

$$H_0 : F_n(x) \in \hat{F}^*(x),$$
$$H_1 : F_n(x) \notin \hat{F}^*(x).$$

The critical values and p-values were calculated using Monte Carlo simulation. The null hypothesis was rejected, when the p-value was lower than the assumed significance level.

The verification of hypothesis is based on three groups of the test statistics. The first group of statistics is based on the supremum, the second on the quadratic functions and the last group consists of the new statistics, specifically designed to test goodness of fit in the upper tail [4].

Supremum version of statistics

The first statistic in this group is the Kolmogorov-Smirnov one, which measures the greatest distance between the empirical and fitted distribution function and usually reaches the maximum value around the median. It's modified version is:

$$KS^* = \max\{KS^{+*}, KS^{-*}\},$$

where:

$$KS^{+*} = \sqrt{n}\sup\{F_n(x) - \hat{F}^*(x)\},$$
$$KS^{-*} = \sqrt{n}\sup\{\hat{F}^*(x) - F_n(x)\}.$$

The second supremum statistic is the Kuiper statistic, which adds the biggest positive and negative differences between the empirical and fitted distribution function. It also reaches the maximum value around the median and is expressed in the form:

$$V^* = KS^{+*} + KS^{-*}.$$

Finally, the third supremum statistic is the Anderson-Darling one, which is calculated by multiplying the Kolmogorov-Smirnov statistic by the weight function. Opposite to the former statistics, this one assigns a higher importance to the lower and upper tails. It's customized version is:

$$AD^* = \sqrt{n}\sup_x \left| \frac{F_n(x) - \hat{F}^*(x)}{\sqrt{\hat{F}^*(x)\left(1 - \hat{F}^*(x)\right)}} \right|.$$

Quadratic version of statistics

This group consists of two statistics: the Cramér-von Mises and Anderson-Darling statistic. Both of them measure the area between empirical and fitted distribution function, but they assign different weights to observations. The Cramér- von Mises weight function is equal to 1 and it is expressed as:

$$W^{2*} = n\int_H^\infty \left(F_n(x) - \hat{F}^*(x)\right)^2 d\hat{F}^*(x).$$

In case of the truncated data, the weight function of the Anderson-Darling statistic equals to $\hat{F}^*(x)(1-\hat{F}^*(x))^{-1}$. The statistic itself is as follows:

$$AD^{2*} = n\int_H^{\infty} \frac{(F_n(x)-\hat{F}^*(x))^2}{\hat{F}^*(x)(1-\hat{F}^*(x))} d\hat{F}^*(x) \cdot$$

New upper tail statistics

Now we will present the two new statistics proposed in [4] to test goodness of fit in the upper tail. Both of them are modifications of the Anderson-Darling statistics, in the supremum and quadratic version accordingly. The authors suggested using the following weight function: $(1-\hat{F}^*(x))^{-1}$. Taking this into account, the supremum and quadratic versions of the new statistics are expressed as:

$$AD_{up} = \sqrt{n}\sup_x \left| \frac{F_n(x)-\hat{F}^*(x)}{1-\hat{F}^*(x)} \right|, \tag{11}$$

$$AD_{up}^2 = n\int_H^{\infty} \frac{(F_n(x)-\hat{F}^*(x))^2}{(1-\hat{F}^*(x))^2} d\hat{F}^*(x) \cdot \tag{12}$$

The derivation of the computational formulas for the quadratic and new statistics presented above can be found in [4].

4. Empirical study

We have tested 19 varied time series, putting a special emphasis on the Polish stock market:

- 5 indices (DAX, NIKKEI, S&P500, WIG, WIG20);
- 5 exchange rates (EURPLN, EURUSD, GBPUSD, USDJPY, USDPLN);
- 5 major companies included in WIG20 index, noted at least since 01.01.2009 (KGHM, PEKAO, PGNIG, PKNORLEN, PKOBP);
- 4 futures (WIG20, gold, coffee and crude oil, abbreviated as FW20, GC_F, KC_F and SC_F).

The time series come from the period of 01.01.2009 - 19.11.2015, so they embrace both peaks and downs. All the calculations have been done using R statistical computing environment. The statistical tests come entirely from the "truncgof" package.

As we are interested only in fitting in the tails, it is more convenient to use the truncated time series. The truncation was made in the following intervals: $[F^{-1}(0), F^{-1}(0.025)]$, $[F^{-1}(0), F^{-1}(0.05)]$, $[F^{-1}(0.95), F^{-1}(1)]$, $[F^{-1}(0.0975), F^{-1}(1)]$. After the truncation, each of 76 time series was subjected to the goodness of fit verification procedure described in details in section 3. The same procedure was repeated for the following distributions: lognormal, the Weibull and Pareto. Altogether we carried out 1596 statistic tests.

The summarizing figures are presented in table 1. Definitely the GPD has the lowest rate of the rejected null hypothesis. In case of this distribution, a number of rejected hypothesis is here approximately the same for all statistics. The rate of rejected hypothesis for lognormal and the Weibull distribution was between 81%-95%, which allows us to conclude, that they are not appropriate for describing the tails in our samples. It is worth noting, that regarding these two distributions, most tests based on statistics AD^2, KS, V and W^2 rejected the null hypothesis. The remaining statistics, that are AD, AD_{up} and AD_{up}^2, gave more differentiated results.

Distribution	Sum	AD	AD^2	KS	V	W^2	AD_{up}	AD_{up}^2
Lognormal	**80**	68	99	91	87	97	51	63
Weibull	**95**	89	100	100	100	100	79	93
GPD	**13**	17	11	12	13	12	13	14

Table 1 Percentage of rejected null hypothesis for all examined intervals

The detailed results of the research with respect to selected intervals are presented in table 2. It is worth noting that for lognormal and the Weibull distributions, a number of rejected null hypothesis for the lower tail is in every case smaller or equal to number of rejected hypothesis for the upper tail. This suggests that there is one way asymmetry between the lower and upper tail. However, regarding two mentioned distributions, the smallest rate of the rejected null hypothesis for a single interval was 75%. For those distributions, p-values of AD, AD^2, KS, V, W^2 and AD_{up}^2 were on average equal to 0.01. P-values of AD_{up} were more dispersed, on average amount-

ed to 0.09 and 0.03 for lognormal and the Weibull distribution accordingly. The more credible results concern the GDP, where we obtain a higher rate of a consistency of the theoretical and empirical distribution in the right tails, especially when the 5% quantile is taken as a threshold. For a higher level of probability the difference between the left and right tail is negligible.

Distribution	Interval	Sum	AD	AD^2	KS	V	W^2	AD_{up}	AD_{up}^2
Lognormal	$[F^{-1}(0), F^{-1}(0.025)]$	**80**	68	99	91	87	97	51	63
	$[F^{-1}(0), F^{-1}(0.05)]$	**75**	63	95	95	95	89	47	42
	$[F^{-1}(0.95), F^{-1}(1)]$	**89**	74	100	100	100	100	58	89
	$[F^{-1}(0.0975), F^{-1}(1)]$	**85**	68	100	100	95	100	58	74
Weibull	$[F^{-1}(0), F^{-1}(0.025)]$	**95**	89	100	100	100	100	79	93
	$[F^{-1}(0), F^{-1}(0.05)]$	**92**	89	100	100	100	100	63	95
	$[F^{-1}(0.95), F^{-1}(1)]$	**99**	100	100	100	100	100	95	100
	$[F^{-1}(0.0975), F^{-1}(1)]$	**97**	95	100	100	100	100	89	95
GPD	$[F^{-1}(0), F^{-1}(0.025)]$	**13**	17	11	12	13	12	13	14
	$[F^{-1}(0), F^{-1}(0.05)]$	**17**	32	16	16	11	16	16	16
	$[F^{-1}(0.95), F^{-1}(1)]$	**9**	5	11	11	11	11	5	11
	$[F^{-1}(0.0975), F^{-1}(1)]$	**14**	16	5	11	16	11	21	16

Table 2 Percentage of rejected null hypothesis – division by intervals

The detailed results of research in respect to the selected groups of assets are presented in table 3. The stocks are the group of assets, that have the highest rate of the rejected null hypothesis for lognormal and the Weibull distribution, although the rate of the rejected hypothesis for other groups is also very high. The values of KS, V and W^2 test statistics for all the distributions are below 10, for AD^2 below 60 and for AD usually below 1000. The dispersion of the remaining AD_{up} and AD_{up}^2 test statistics is visibly higher, as they reach the level of a several thousand for lognormal, the Weibull and Pareto distributions. In the case of the GPD, we obtain definitely the highest rate of the rejected null hypothesis in the case of exchange rates and the fewest rate for the indices.

Distribution	Type of assets	Sum	AD	AD^2	KS	V	W^2	AD_{up}	AD_{up}^2
Lognormal	Indices	**81**	60	100	95	100	100	45	70
	exchange rates	**78**	70	100	85	80	100	50	60
	Stocks	**87**	90	100	90	80	95	80	75
	Futures	**70**	50	94	94	88	94	25	44
Weibull	Indices	**92**	85	100	100	100	100	70	90
	exchange rates	**94**	85	100	100	100	100	80	90
	Stocks	**99**	100	100	100	100	100	90	100
	Futures	**94**	88	100	100	100	100	75	94
GPD	Indices	**6**	10	5	5	5	10	5	5
	exchange rates	**25**	20	25	30	25	20	25	30
	Stocks	**11**	30	5	5	10	10	10	10
	Futures	**9**	6	6	6	13	6	13	13

Table 3 Percentage of rejected null hypothesis – division by type of assets

5. Summary

An appropriate description of the tails of returns distribution has an important role for the measurement of the extreme risk not yet sampled in the past. It can determine a proper risk management procedure. Depending on a distribution a risk manager chooses, he or she implicitly assumes a very different tail behaviour. The best way to estimate a risk at a high probability level is to use only data coming from the tails instead of a whole distribution sample. The assessment of the quality of an estimation in the tails requires sophisticated tools, like goodness-of-fit tests for the truncated distribution. The paper takes into account several tests used in the empirical study. We verified three types of distributions: lognormal, the Weibull and Generalized Pareto.

Definitely Generalized Pareto Distribution is the distribution that captures the tails the most accurately. Only in thirteen percent of the considered cases, the theoretical model for the tails was inconsistent with the empirical distribution. Besides, in the case of the GPD, a number of the rejected null hypothesis was approximately the same for all the statistics. The remaining two distributions seemed to be inappropriate because the four statistics: AD^2, KS, V, W^2, almost always rejected the null hypothesis of the consistency of the theoretical and empirical distributions. The reason can be the fat tails which cannot be captured by the two distributions. The higher rate of the consistency of the theoretical and empirical distribution we obtained in the right tails, especially when the 5% quantile was taken as a threshold. It can suggest that there is a one way asymmetry between the lower and upper tail, which can cause difficulties to model the left hand side of distribution. On the other hand the GPD is a distribution for all types of domain of attraction. For quantile of 2.5% the differences between the statistic results for the left and right tails are negligible. We also found a significant differences in precision of an estimation among various types of assets. The exchange rates are the group of assets, that have the highest rate of rejected null hypothesis for the GPD, and the lowest rate, approximately four times less, we obtained for the stock indices. Such results were confirmed by all the goodness of fit tests.

References

[1] Bradley, B., Taqqu, M. S.: Financial Risk and Heavy Tails. In: *Handbook of Heavy-Tailed Distributions in Finance* (Rachev S.T., eds.). Amsterdam, 2003, 35–103.

[2] Barndorff-Nielsen, O.: Exponentially decreasing distributions for the logarithm of particle size'. *Proceedings of the Royal Society of London*. Series A, Mathematical and Physical Sciences **353**, 1977, 401-419.

[3] Black, F., Scholes, M.:, The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* **81**(3), 1973, 637–654.

[4] Chernobai, A., Rachev, S.T., Fabozzi, F.J.: Composite goodness-of-fit tests for left-truncated loss samples. *Handbook of Financial Econometrics and Statistics*, Springer, 2014, 575-596.

[5] Dowd, K.: *Measuring Market Risk*. John Willey & Sons Ltd, West Sussex, 2005.

[6] Eberlein, E., Keller, U.: Hyperbolic Distributions in Finance. *Bernoulli* 1, 1995, 281-299.

[7] Eberlein, E., Keller, U., Prause, K.: New Insights into Smile, Mispricing and Value at Risk: The Hyperbolic Model. *Journal of Business* **71**(3), 1998, 371–405.

[8] Embrechts, P., Sidney, I., Resnick, S.I., Samorodnitsky, G., Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal* **3**(2), 1999.

[9] Fama, E.: The Behavior of Stock Market Prices. *Journal of Business* **38**, 1965, 34-105.

[10] Fielitz, B.D.: Stationary or random data: Some implications for the distribution of stock price changes. *Journal of Financial and Quantitative Analysis* **6**, 1971, 1025-1034.

[11] Goldmann, C., Klar, B., Meintanis, S.G.: Data transformations and goodness-of-fit tests for type-II right censored samples. *Metrika* **78**, 2015, 59-83.

[12] Haas, M., Pigorsch, C.: Financial Economics, Fat-tailed Distributions. *Encyclopedia of Complexity and Systems Science* **4**, 2007.

[13] Kabasinskas, A., Rachev, S. T., Sakalauskas, L., Sun, W., Belovas, I.: Alpha-Stable Paradigm in Financial Markets. *Journal of Computational Analysis and Applications* **11**(4) , 2009, 641-668.

[14] Küchler, U., Neumann, K., Sørensen, M., Streller, A.: Stock Returns and Hyperbolic Distributions. *Mathematical and Computer Modelling* **29**,1999, 1–15.

[15] Lévy, P.: *Calculation des Probabilités*. Gauthier-Villars, Paris, 1995.

[16] Mandelbrot, B.: The Variation of Certain Speculative Prices. *Journal of Business* **36**, 1963, 394-419.

[17] Merton, R. C.: Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science* **4**(1), 1973, 141–183.

[18] Mitzenmacher, M.: A brief history of generative models for power law and lognormal distributions. *Internet Mathematics* **1** (2), 2004, 226-251.

[19] Rachev, S., Mittnik, S.: *Stable Paretian Models in Finance*. Wiley, Chichester, 2000.

[20] Rachev, S.: *Handbook of Heavy-tailed Distributions in Finance*, North Holland, 2003.

[21] Raible, S.: *Lévy Process in Finance: Theory, Numerics, and Empirical Facts*. Ph.D. thesis, Freiburg University, 1998.

[22] Weron, R.: Computationally Intensive Value at Risk Calculations. In: *Handbook of Computational Statistics* (Gentle, J. E., Haerdle, W., Mori T., eds.). Springer, Berlin, 2004, 911-950.

Backtesting Value-at-Risk for multiple risk levels: a lexicographic ordering approach

Marcin Fałdziński[1]

Abstract. The main aim of this analysis is to provide a new procedure for backtesting in the case of the Value-at-Risk (VaR) based on multiple coverage probabilities at once, i.e. a multivariate approach. Our procedure is based on the lexicographical ordering of the coverage probabilities. Backtesting has a great meaning due to the fact that differentiating risk models which improperly estimate market risk constitute one of the crucial challenges in quantitative risk management. It is assumed, that a given model should provide a good accuracy of VaRs for any coverage probabilities or a combination of them. Conventionally, VaR is backtested for one coverage probability which may result in choosing a model that is not accurate in a multivariate case. In this paper, we present and test a new procedure of backtesting Value-at-Risk in a multivariate setting using two tests proposed by [22] and [17]. In the empirical study, various methods for market risk estimation were used: ARMA-GARCH models, ARMA-GARCH with Peaks over Threshold (POT) method, historical simulations, and The RiskMetrics.

Keywords: backtesting, Value-at-Risk, quantitative risk management

JEL Classification: C58, G17
AMS Classification: 90C15

1 Introduction

Value-at-Risk (hereafter VaR) is one of the most commonly used measures for market risk. Basel Committee on Banking Supervision implemented Value-at-Risk as a standard market risk measure used by financial institutions and supervising authorities. The 1995 Market Risk Amendment to Basel I Accord provides a spectrum of approaches for determining market risk capital requirements. According to the internal model approach financial institutions are allowed to estimate the capital requirement for market risk using their internal models. The 2004 Basel II Accord was created to reward institutions with superior risk management systems. In cases where internal models led to a greater number of violations (losses higher than VaR) than could be expected, given the coverage level, banks are required to hold a higher level of capital. The greater number of violations the higher level of capital to hold, so the need of accurate VaR model is considerable. Value-at-Risk is being used not only as market risk measure, but as measure of financial contagion or spill-over effect [3], [12], [13], [14], [16], [21]. Backtesting is a set of specifically design statistical methods to test the performance of VaR. We cannot use the standard evaluation methods (for instance MAPE) because ex-post VaR is not observable.

2 Research Methodology

Formally, r_t is the return of an asset or a portfolio of assets at time *t*, then Value-at-Risk for a $\alpha\%$ coverage level, conditional to an information set Ω_{t-1}, available at time $t-1$ is defined as $P\left\{r_t \le -VaR_{t|t-1}(\alpha) \mid \Omega_{t-1}\right\} = \alpha$ $\alpha \in (0,1)$. Typically, the coverage level (or coverage probability) is chosen to be either 5% or 1% or 10%. The hit variable (or violation variable) associated with the ex-post observation of a $VaR_{t|t-1}(\alpha)$ at time *t*, denoted $I_t(\alpha)$ is defined as:

$$I_t(\alpha) = \begin{cases} 1 & \text{if } r_t \le -VaR_{t|t-1}(\alpha) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

[5] shows that in order to assess VaR validity it is possible to test whether the hit sequence $\{I_t\}_{t=1}^{T}$ follows two conditions:

1. Unconditional coverage (UC) [5], [18]

[1] Nicolaus Copernicus University/ Department of Econometrics and Statistics, ul. Gagarina 13a, Toruń, Poland. marf@umk.pl.

$P[I_t(\alpha)=1]=E[I_t(\alpha)]=\alpha$ or equivalently $I_t(\alpha) \underset{i.i.d.}{\sim} Bernoulli(\alpha)$

2. Independence property (IND) [5]

the variable $I_t(\alpha)$ has to be independent from the variable $I_{t-k}(\alpha) \forall k \neq 0$.

These two conditions are necessary but not sufficient of VaR definition. The gap between these conditions is known as model risk ([10]). There are many tests which exploit the violation sequence $I_t(\alpha)$ e.g. [1], [2], [4], [5], [6], [8], [9], [10] and [23]. Unfortunately, all these tests assume that we check UC and IND for a single coverage level. Worth mentioning is that a given VaR model should provide good accuracy for any coverage levels. Let $A=\{\alpha_1,...,\alpha_m\}$ is a discrete set of *m* different coverage levels. By the multivariate approach, we mean that the VaR validity is tested based on the discrete set of *m* different coverage levels at the same time, so we focus on the left tail of the distribution instead of on the one point only. [23] derived the multivariate unconditional coverage test (MUC) as the multivariate case of the unconditional coverage test of [19] approach. We focus on VaR with *m* different coverage levels, indexed *i* in descending order, $\alpha_1 > \alpha_2 > ... > \alpha_m$. Now we build a new violation indicator:

$$I_{i,t}=\begin{cases}1 & \text{if } r_t \leq -VaR_{t|t-1}(\alpha_i) \\ 0 & \text{if } r_t > -VaR_{t|t-1}(\alpha_i)\end{cases} \tag{2}$$

for $i=1,...,m$ with $J_{i,t}=I_{i,t}-I_{i+1,t}$ .The $J_{i,t}$ Bernoulli random variables equal one with probability $\theta_i=\alpha_i-\alpha_{i+1}$ when $i>0$ and $J_{0,t}=0$ with probability $\theta_{0,t}=1-\alpha_1$. Let $\theta=(\theta_1,...,\theta_m)'$ be the m-dimensional vector of θ_i. The statistic is a likelihood ratio test given by:

$$LR_{MUC}=2\left[n_0\ln\left(\mathbf{1}-\mathbf{1}'\hat{\boldsymbol{\theta}}\right)+\sum_{i=1}^{m}n_i\ln\hat{\theta}_i\right]-2\left[n_0\ln\left(\mathbf{1}-\mathbf{1}'\boldsymbol{\theta}\right)+\sum_{i=1}^{m}n_i\ln\theta_i\right] \tag{3}$$

where $n_i=\sum_{t=1}^{T}J_{i,t}$ and $\hat{\theta}_i=(1/T)\sum_{t=1}^{T}J_{i,t}$ are the maximum likelihood estimates. The LR_{MUC} statistic is asymptotically χ^2 with *m* degrees of freedom. When $m=1$ then the multivariate test boils down to the univariate unconditional coverage test of [19]. [17] proposed a multivariate portmanteau test of the absence of autocorrelation of violations. Let, $h_t(\alpha)$ is associated the value $1-\alpha$ in case of violations and $-\alpha$ otherwise, which can be written as: $h_t(\alpha)=I_t(\alpha)-\alpha$. Now let, $H_t=[h_t(\alpha_1),...,h_t(\alpha_m)]$ is the vector grouping the violation sequences associated with these *m* coverage levels at time *t*. The test consists in the joint nullity of the autocorrelations order 1 to *L* for the vector H_t, $H_0: E[h_t(\alpha_i)h_{t-l}(\alpha_j)]=0$, for the lag $l=1,...,L$ and coverage levels $(\alpha_i,\alpha_j)\in A$. [17] use a multivariate portmanteau statistics to construct the test statistics. If $\boldsymbol{\Omega}_l$ is the matrix of empirical covariance associated to a vector H_t such as: $\boldsymbol{\Omega}_l=\sum_{t=l+1}^{L}H_tH_{t-l}^{'}$ for the set of lags $l=1,..,L$. The statistic for the test is given by:

$$Q_m(L)=T^2\sum_{l=1}^{L}(T-l)^{-1}Diag\left(\boldsymbol{\Omega}_l^{'}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Omega}_l\boldsymbol{\Omega}_0^{-1}\right)^{'}\mathbf{1} \tag{4}$$

The $Q_m(L)$ statistic is asymptotically χ^2 with Tm^2 degrees of freedom.

Both [17] and [22] show that these multivariate tests are better than univariate ones in terms of accuracy or ability to reject the misspecified VaR model. In order to further improve the multivariate approach we propose to use the lexicographic ordering of the assumed coverage probabilities, i.e., we should test all possible combinations of *m* different coverage levels starting from all one element sets, then all two elements sets, all three elements sets and so on up to last *m* elements set of coverage levels. For instance, for a rather small set of three coverage levels $(\alpha_1=0.01,\alpha_2=0.05,\alpha_3=0.1)$, we would get three one-element sets i.e. $(\alpha_1),(\alpha_2),(\alpha_3)$, all two-elements sets i.e. $(\alpha_1,\alpha_2),(\alpha_1,\alpha_3),(\alpha_2,\alpha_3)$ and the last set of all three elements $(\alpha_1,\alpha_2,\alpha_3)$ which gives in total 7 combinations. In short, the lexicographic ordering means that we order coverage levels in the testing set A from the lowest coverage level to the highest. The basic idea behind this approach is that we utilize the larger information set in

order to evaluate VaR model. Secondly, and more importantly, one model can have better accuracy at lower coverage levels, but worse on higher coverage levels. This is rather evident from the practical and theoretical point of view ([13], [15], [16] and [18]). In the case of the set of all three coverage levels $(\alpha_1, \alpha_2, \alpha_3)$ only, it could be quite hard to detect cases when for, let say, $\alpha_1 = 0.01$ the null hypothesis should be rejected. This would be even harder when the set of coverage levels A comprises of 5, 10 or more coverage probabilities. The reason for that is the LR_{MUC} and $Q_m(L)$ statistics take into account all coverage levels at once with the same weights. In order to prevent it from happening, we should test all possible combinations of coverage levels, so that we can find at which coverage level or levels there is an indication to reject the null hypothesis.

3 Empirical Research

In the research daily prices from 3rd January 2000 until 3rd January 2012 of stock market indices were used, namely S&P 500, FTSE 100, DAX, CAC 40 and WIG. This period comprises of 3000 daily logarithmic returns. In the first step of the research we estimated five competing models:

1. ARMA(p,q)-GARCH(1,1) applying Peaks over Threshold (POT) method ([21]) with t-distributon and normal distribution (hereafter ARMA-GARCH-POT t-dist.). The lag order for ARMA(p,q) was selected based on BIC criterion where the maximum lag order was set to 5.
2. ARMA(p,q)-GARCH(1,1) with t-distributon and normal distribution (in short ARMA-GARCH t-dist.). The lag order for ARMA(p,q) was selected based on BIC criterion where the maximum lag order was set to 5.
3. Historical simulation defined as p-percentile of the returns distribution.
4. The RiskMetrics method where the conditional variance h_t is modeled by: $h_{t+1} = \lambda h_t + (1-\lambda)R_t^2$ with $\lambda = 0.94$.
5. GARCH(1,1) model with a normal distribution as a benchmark.

For the estimation of the parameters, we have applied a moving window of 1000 returns which were re-estimated after adding one observation to the sample and then computing VaR. The whole process was done after reaching the end of the sample, which gave in total 2000 VaR estimates for comparison. We decided to use a discrete set of ten different coverage levels $A = \{0.01, 0.02, ..., 0.09, 0.1\}$, which gives a total of 1023 combinations for the lexicographic ordering. For $Q_m(L)$ test we used five lag orders $L = \{1, 2, ..., 5\}$. The maximum likelihood estimator has been used to estimate parameters of the aforementioned models.

Table 1 presents results for multivariate unconditional coverage test LR_{MUC} and a multivariate portmanteau test

Model/Method	LR_{MUC}	p-value	$Q_m(L=1)$	p-value	$Q_m(L=2)$	p-value
ARMA-GARCH t-dist.	9.3656	0.4978	102.3613	0.4157	202.4048	0.4392
ARMA-GARCH normal dist.	20.7848	**0.0226**	88.0185	0.7984	174.2861	0.9052
GARCH normal dist.	21.3676	**0.0187**	99.4321	0.4972	199.0753	0.5052
Historical simulation	16.2858	0.0917	116.7623	0.1207	241.2649	**0.0245**
RiskMetrics	43.1850	**0.0000**	117.0931	0.1165	238.2215	**0.0333**
ARMA-GARCH-POT t-dist.	11.7807	0.3000	117.5645	0.1108	229.5774	0.0743
ARMA-GARCH-POT normal dist.	12.3950	0.2595	122.8139	0.0605	257.5782	**0.0037**
Model/Method	$Q_m(L=3)$	p-value	$Q_m(L=4)$	p-value	$Q_m(L=5)$	p-value
ARMA-GARCH t-dist.	308.1849	0.3601	375.6601	0.80371	474.8411	0.7847
ARMA-GARCH normal dist.	281.5595	0.7708	360.5576	0.9221	460.8198	0.8947
GARCH normal dist.	303.3148	0.4356	366.3149	0.8854	497.0537	0.5288
Historical simulation	403.6516	**0.0001**	519.1452	**0.0001**	657.9405	**0.0000**
RiskMetrics	365.3573	**0.0058**	453.4141	**0.0333**	578.4748	**0.0086**
ARMA-GARCH-POT t-dist.	372.8909	**0.0026**	470.5870	**0.0085**	573.1476	**0.0129**
ARMA-GARCH-POT normal dist.	409.2781	**0.0000**	523.8963	**0.0000**	622.0574	**0.0002**

Table 1 Multivariate LR_{MUC} test and $Q_m(L)$ test results for FTSE 100 assuming $A = \{0.01, 0.02, ..., 0.09, 0.1\}$

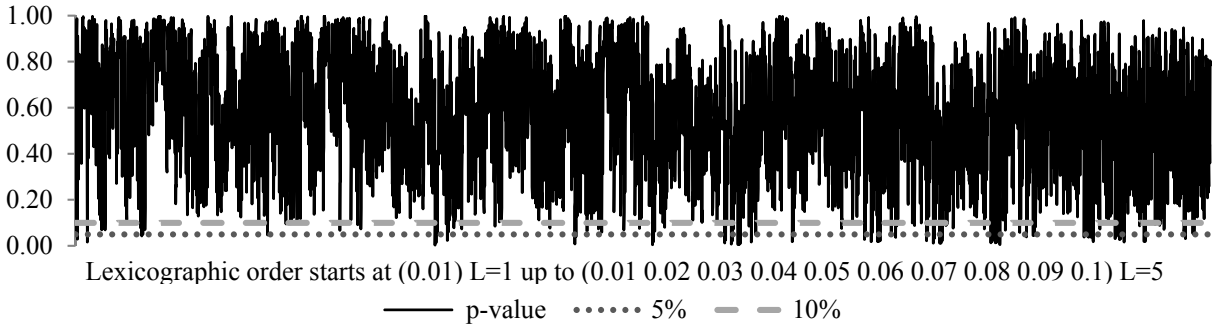


Figure 1 p-values for $Q_m(L)$ statistics from ARMA-GARCH t-dist. for FTSE 100 index

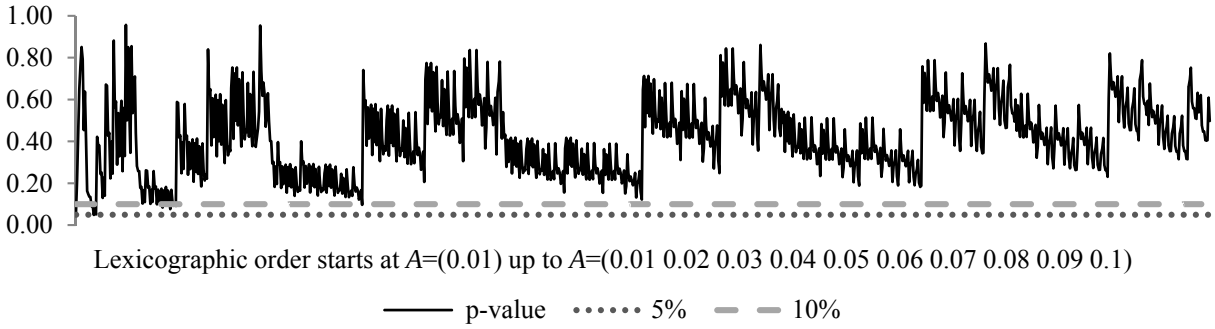


Figure 2 p-values for LR_{MUC} statistics from ARMA-GARCH t-dist. for FTSE 100 index

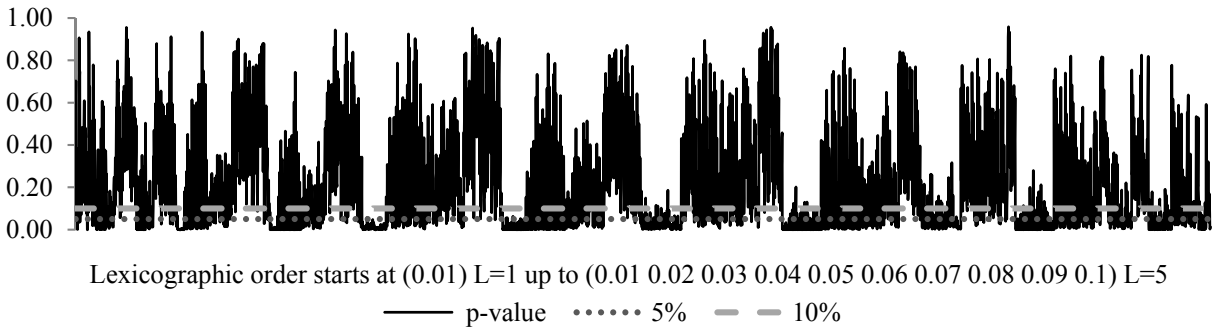


Figure 3 p-values for $Q_m(L)$ statistics from ARMA-GARCH-POT t-dist. for FTSE 100 index

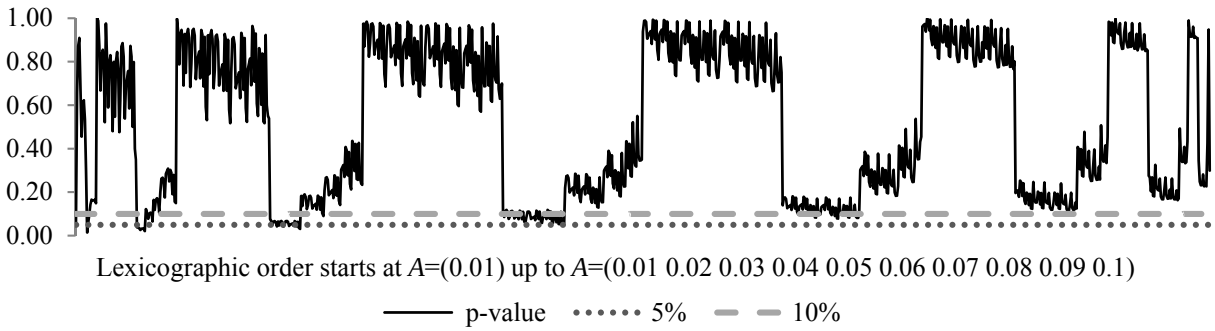


Figure 4 p-values for LR_{MUC} statistics from ARMA-GARCH-POT t-dist. for FTSE 100 index

Additionally, for comparison reasons the standard test for unconditional coverage LR_{UC} [19] and independence property test LR_{IND} [5] were computed. P-values have been obtained from asymptotic distributions, but this can be improved upon. One can apply [7] method to simulate the distribution of the test statistics (3) and (4) instead of using the asymptotic distribution. In order to save space, we present only part of the results (all the results are available upon request) of the absence of autocorrelation of violations $Q_m(L)$ for FTSE 100 index while assuming

a whole set of coverage levels $A = \{0.01, 0.02, ..., 0.09, 0.1\}$. As we can see for four models (ARMA-GARCH t-dist., historical simulation, ARMA-GARCH-POT t-dist. and ARMA-GARCH-POT normal dist.) we do not reject the null hypothesis at the 5% significance level, so the unconditional coverage condition is met. In the case of the independence property, the null hypothesis is not rejected (5% significance level) for all models at lag L=1, but at lags L=3,4,5 we reject the null for historical simulations, The RiskMetrics, ARMA-GARCH-POT normal dist. and ARMA-GARCH-POT t-dist. It is not so obvious which model is the best. Figure 1 and figure 2 show p-values for LR_{MUC} and $Q_m(L)$ when the lexicographic ordering is applied for ARMA-GARCH t-dist. model. Figure 3 and figure 4 show p-values for LR_{MUC} and $Q_m(L)$ when the lexicographic ordering is applied for ARMA-GARCH-POT t-dist. model. It is easily visible that there are some combinations of coverage levels which lead to rejection of the null for both tests LR_{MUC} and $Q_m(L)$ particularly for ARMA-GARCH-POT t-dist. model. Specifically, high coverage levels lead to rejection of the null. In order to have a better grasp of the situation, table 2 give percentages of rejection of the null hypothesis (5% significance level) for the multivariate unconditional test and test for absence of autocorrelation for violations Simplifying, we can say that the higher the percentage level, the worse accuracy of VaR model. In the case of the FTSE 100, ARMA-GARCH t-dist. seems to be the best model for VaR estimation, where we have 0.1% and 1.33% of cases when the null is rejected for LR_{MUC} and $Q_m(L)$ test respectively. For all indices, we can more easily choose a model, which has the best accuracy considering two aforementioned tests. Generally, we can say that ARMA-GARCH-POT t-dist. model and ARMA-GARCH model with t-distribution give the most accurate VaR estimates.

LR_{MUC}					
Model/Method	WIG	DAX	CAC 40	FTSE 100	S&P 500
ARMA-GARCH t-dist	59.43%	0.00%	22.80%	0.10%	6.06%
ARMA-GARCH normal dist	98.92%	75.37%	44.87%	65.98%	96.38%
GARCH normal dist	80.55%	66.28%	94.62%	79.37%	94.81%
Historical simulation	50.05%	30.99%	98.04%	32.26%	52.98%
RiskMetrics	66.76%	97.07%	56.60%	96.77%	97.07%
ARMA-GARCH-POT t-dist	0.29%	0.00%	0.00%	1.66%	0.29%
ARMA-GARCH-POT normal dist	0.20%	0.00%	0.10%	56.13%	68.91%
$Q_m(L)$ $L = 1,...,5$					
Model/Method	WIG	DAX	CAC 40	FTSE 100	S&P 500
ARMA-GARCH t-dist	94.15%	44.11%	23.75%	1.33%	33.98%
ARMA-GARCH normal dist	98.18%	58.22%	0.86%	3.03%	20.10%
GARCH normal dist	93.04%	51.48%	18.71%	4.22%	21.09%
Historical simulation	99.00%	99.57%	98.04%	88.58%	61.31%
RiskMetrics	62.85%	75.62%	29.31%	54.68%	8.02%
ARMA-GARCH-POT t-dist	67.88%	75.52%	34.49%	36.72%	15.11%
ARMA-GARCH-POT normal dist	59.77%	81.51%	51.95%	0.20%	43.28%

Table 2 Percentage when the null hypothesis is rejected at 5% significance level

4 Conclusions

This paper focuses on the issue of backtesting Value-at-Risk models. We argue that using a multivariate approach with the lexicographic ordering can better differentiate among VaR models. The multivariate approach itself is an improvement of the existing methods due to the fact that we concentrate our attention at the left tail of the distribution instead of some points of the distribution, i.e. we use a discrete set of the coverage levels instead of one coverage level. The proposed lexicographic ordering is based on the idea that we test all possible combinations of a set of the coverage levels, which enable us to better distinguish cases when some coverage levels lead to worse properties of VaR model. This approach could be helpful in finding a model which gives VaR with desired properties.

References

[1] Berkowitz J., Christoffersen P. F. and Pelletier D.: Evaluating Value-at-Risk models with desk-level data, (2005), *Working Paper*, University of Houston.

[2] Candelon B., Colletaz G., Hurlin C. and Tokpavi S.: Backtesting Value-at-Risk: A GMM Duration-based Test, *Journal of Financial Econometrics* **9** (2011), 314–343.

[3] Candelon, B., Joëts, M. and Tokpavi, S.: Testing for Granger Causality in Distribution Tails: An Application to Oil Markets Integration, *Economic Modelling* **31** (2013), 276–285.

[4] Colletaz G., Hurlin Ch. and Perignon Ch.: The Risk Map A new tool for validating risk models, *Journal of Banking and Finance* **37** (2013), 3843-3854

[5] Christoffersen P.F.: Evaluating Interval Forecasting, *International Economic Review* **39** (1998), 841-862.

[6] Danciulescu, C.: Backtesting Value-at-Risk Models: A Multivariate Approach (April 3, 2010). Center for Applied Economics & Policy Research Working Paper No. 004-2010. Retrieved from https://ssrn.com/abstract=1591049.

[7] Dufour J. M.: Monte Carlo Tests with Nuisance Parameters: a General Approach to Finite Sample Inference and Nonstandard Asymptotics, *Journal of Econometrics* **27** (2006), no.2, 443-477.

[8] Engle R. F. and Manganelli S.: CAViaR: Conditional Autoregressive Value-at-Risk by Regression Quantiles. *Journal of Business and Economic Statistics* **22** (2004), 367-381.

[9] Campbell S. D.: A Review of Backtesting and Backtesting Procedures, *Working Paper*, (2005), Finance and Economics Discussion Series.

[10] Escanciano J.C. and Olmo J.: Backtesting Parametric Value-at-Risk With Estimation Risk, *Journal of Business & Economic Statistics* **28** (2010), No. 1, 36-51.

[11] Fałdziński M.: On The Empirical Importance Of The Spectral Risk Measure With Extreme Value Theory Approach, Financial Markets Principles of Modelling Forecasting and Decision-Making, *FindEcon*, 2011, 73-86

[12] Fałdziński M., Balcerzak A.P., Meluzín T., Pietrzak M.B. and Zinecker M.: Cointegration of interdependencies among capital markets of chosen Visegrad countries and Germany. In*: Proceedings of the 34th International Conference Mathematical Methods in Economics 2016* (Kocourek, A. and Vavroušek, M., eds.) Technical University of Liberec, Liberec, 2016, 189-194.

[13] Fałdzinski M. and Osińska M.: Volatility Estimators in Econometric Analysis of Risk Transfer on Capital Markets, *Dynamic Econometric Models* **16** (2016), 21-35.

[14] Fałdziński M., Osinska M. and Zdanowicz T.: Detecting Risk Transfer in Financial Markets using Different Risk Measures, *Central European Journal of Economic Modelling and Econometrics* **1** (2012), vol. 4, 45-64.

[15] Harmantzis F.C., Miao L. and Chien Y.: Empirical study of value-at-risk and expected shortfall models with heavy tails, *Journal of Risk Finance* **2** (2006), Vol. 7, 117-126.

[16] Hong Y., Liu Y. and Wang S.: Granger causality in risk and detection of extreme risk spillover between financial markets, *Journal of Econometrics* **2** (2009), 150, 271–287.

[17] Hurlin Ch. and Tokpavi S.: Backtesting Value-at-Risk Accuracy: A New Simple Test, *Journal of Risk* **9** (2007), 19-37,

[18] Kuester K., Mittnik S. and Paolella M.S.: Value-at-Risk Prediction: A Comparison of Alternative Strategies, *Journal of Financial Econometrics* **1** (2006), vol. 4, 53-89.

[19] Kupiec P.: Techniques for Verifying the Accuracy of Risk Measurement Models, *Journal of Derivatives* **3** (1995), 73-84.

[20] McNeil J.A.,and Frey F.: Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach, *Journal of Empirical Finance* **7** (2000), 271–300.

[21] Osińska M., Fałdziński M. and Zdanowicz T., Econometric Analysis of the Risk Transfer in Capital Markets. The Case of China, Argumenta Oeconomica, 2 (2012), 139-164.

[22] Perignon Ch. and Smith D.: A New Approach to Comparing VaR Estimation Methods, *Journal of Derivatives* **16** (2008), 54-66.

[23] Ziggel D., Berens T., Weiss G.N. and Wied D.: A New Set of Improved Value-at-Risk Backtests., *Journal of Banking and Finance* **48** (2014), 29-41.

Models of Equilibrium on Network Industries Market in Context of Influence of Regulated Prices in Slovakia

Eleonora Fendeková[1], Michal Fendek[2]

Abstract. In Slovakia in the recent period of time we have seen an intensive discussion concerning the effectiveness of the price regulation methods in Network Industries and the objectivity and the labor social effectiveness of The Regulatory Office for Network Industries, as well. In the paper we try to present qualified professional view on this problems.
The aim of the paper is to examine the equilibrium conditions in the market of network industries. This issue bears an increased attention under conditions of economic influence of network industries in the national economies since the second half of the twenties of the last century. This tendency is especially visible particularly in the context of economic globalization and increased significance of the supranational entities.
The existence of pure monopoly in network industries increases the role of regulation mechanisms in connection with objectification and increase in their social effectiveness. The objective of regulation mechanisms is to find an appropriate proportion between price and product supply of network industry.
The conditions for equilibrium of network industries and methods of their regulations will be examined in the paper. The stress will be laid on the regulation on the performance basis – Performance based Regulation, regulation on the base of returns – Rate of Return Regulation and also on the price regulation on the basis Averch-Johnson model. Attention will be paid to the ways of calculation reasonable profit in regulated industries and assessment of effectiveness and influence of regulation measurements on Slovak network industries.
Keywords: Competitive environment, network industries, regulated prices, reasonable profit in regulated industries, quantitative analysis.

JEL Classification: D43, L13, L43, L44
AMS Classification: 91B16, 91B24

1 Introduction

For current optimization pricing strategies in the network industries, it is typical that traditional pricing mechanisms corresponding to the now-classifying scheme of a perfectly competitive market environment are replaced by adequate mechanisms for monopoly and oligopolistic market structures, which are enriched by analytical regulation schemes in case of network operators.

Under the conditions of the Slovak Republic network industries are regulated. Their name reflects the fact that business in these sectors is only possible through networks, the system of technical means through which entrepreneurs in these sectors meeting the attributes of natural monopolies deliver goods or provide services, for example, electricity network, gas network, telecommunication network, etc. In Blum, Müller, and Weiske von Gabler [2] we can see, that This modern segment of microeconomic analysis, which examines the patterns of equilibrium between supply and demand in transnational markets using the apparatus of optimization methods, has become a separate discipline of economic research in the English scientific literature known as the Industrial Organization in recent decades, respectively in the German literature as Industrieökonomik.

Pepall, Richards and Norman [9] show, that a monopoly, by virtue of its exclusive position on the market as the sole supplier of a product or service on the relevant market, may set a market price and offer a volume of production that will enable it to maximize profits beyond what a competitor's business may have in that context and undesirable social effects.

[1] University of Economics Bratislava, Department of Business Economics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, e-mail: nfendek@dec.euba.sk.
[2] University of Economics Bratislava, Department of Operations Research and Econometrics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, e-mail: fendek@dec.euba.sk.

In Bodea and Fergusson [1], we can see, that the issue of price regulation in the network sector market is currently a particularly topical issue, as the issues of reasonable profit and objectively justified costs of network operators now resonate relatively intensively in the professional public. It should be accepted that the traditional pricing mechanisms corresponding to the now-classifying scheme of a perfectly competitive market environment are replaced by adequate mechanisms for monopoly and oligopolistic market structures, which are even enriched by regulatory analytical schemes for network operators.

For the regulation of network operators who represent, in terms of their position on the market, monopoly or natural monopolies, the state creates a so-called Regulator, whose role is ultimately the state's mandate to create a legislative environment and regulatory mechanisms that will ensure a market equilibrium for the commodity underlying the guarantee of a reasonable profit for the regulated entity.

Regulatory authorities in connection with the fulfillment of their basic mission, which is the factual and price regulation of the business in the regulated activities of selected sectors, is in place before the effective solution of the two tasks is achieved:

a) the creation of a functioning and, at the same time, it is necessary to emphasize the competitive market environment with the products of the network industries by applying standard regulatory mechanisms, especially in the context of the successful establishment of Slovakia in the European Union space and the gradual adaptation of Slovakia to the conditions in the energy markets A united Europe;

b) to prepare an analytical apparatus for price regulation of network industries, which would guarantee the effective development of regulated entities in Slovakia.

2 Legislative tools for network industries regulation and effects of regulation in Slovak republic

For the regulation of entities in selected sectors, which represent, in terms of their position on the market, monopolies or, Natural monopolies, the state creates a so-called Regulator, whose role is to create a legislative environment and regulatory mechanisms that will ensure market equilibrium for the monitored commodities in guaranteeing a reasonable profit for the regulated entity. These industries include, in particular, network industries whose subjects are often monopolistic in nature, and the state is therefore keen to regulate their activities.

Important part of market with electricity, gas, and other network industries products restructuring in Slovak republic was the creation of a new regulatory framework. In august 2001 The Regulatory Office for Network Industries (hereinafter only "Regulatory Office" or "RONI") has been founded. Its role is to issue licenses, regulate prices and quality norms for network industries products. In the years 2001 to 2005 the Regulatory Office defined the price regulation regime and worked out a framework of new price regulation rules for network industries products price setting.

The result of The Ministry of Economy, Antimonopoly Office of the Slovak Republic and numerous legal, economics and regulation experts effort is the Act No. 276/2001 Coll. on Regulation of network industries, which has been passed by National Council of the Slovak Republic at 14th June 2001 and came into force at 1st August 2001. Force of articles regarding price regulation has been suspended until 1st January 2003. On the date when law came into force the Regulatory Office for Network Industries has been found, as a state administration body.

Regulatory Office for Network Industries has been found under the rule of 276/2001 Act, which has been modified and amended by Act No. 107/2007 Coll. Act No. 107/2007 Coll. is clearly defining price regulatory state administrative bodies in network industries which are:

a) Regulatory Office for Network Industries,
b) The Regulatory Board.

The basic legal regulation that regulates the network industries in Slovakia is Act no. 276/2001 Z.z. Regulation on network industries. According to this law, the network industries mean:
- Production, transmission, distribution and supply of electricity and related services,
- The production, transport, distribution, storage and supply of gas and related activities
- Heat generation and distribution,
- Water management and collective drainage of public sewage.

The regulation is defined in the applicable law as defined as ensuring non-discriminatory and transparent performance of activities in network industries, the application of regulatory measures aimed at reducing the risks of breach of competition rules by abusing a dominant position on the goods and services market in order to protect the rights and obligations of eligible customer's households.

Another important change, which was brought by Regulation Act and which is under the competence of RONI was providing a third party access to the networks. The owner of transit (distribution) network license is obliged by contract to provide transfer and distribution of electricity or transport and distribution of gas for reward for eligible customer, who will meet technical conditions of this network.

Regulatory office oversees following of obligation to provide a third party with access to the networks and adjudicates disputes. Obligation to provide a third party with network access is also related to full opening of the internal energy market and to consumer option to freely choose energy supplier. According to this, RONI regulations are limited to activities with natural monopoly character only, that is, prices or tariffs set by the office will not include any unregulated items.

RONI is carrying out regulation through the use of standard, in developed economies proven methods. Details about price regulation process in network industries and determination of scope of justified costs and reasonable profit are enacted by the office in office decrees, which are published in Collection of Acts, in Office bulletin and on the website of the office.
Efficient functioning of commodity and service market in network industries in conditions of Slovak republic can be insured only by quality and potent regulatory policy in the context with the Slovak Republic law and respecting the adopted legal acts of European Community and European Union. All consumers, with emphasis on "vulnerable" groups, mainly households, must have an opportunity to make use of high level consumer protection, referring mainly to households, while maintaining adequate guarantees and suitable incentives for regulated subjects.

3 Microeconomics models of network industries price regulation

Under the conditions of the Slovak Republic regulation is subject to the so network industries. Business in these sectors is only possible through networks, i.e. systems of technical means through which entrepreneurs in these sectors supply goods or provide services, electricity network, gas network, and so on. As a result of its exclusive position, a monopoly can set a market price and offer a volume of production that will enable it to maximize profits beyond the possibilities of a competing firm.

Therefore, it is desirable to have the tools to redistribute this gain of the monopoly so that part of it supports the whole-party goals of the development of the economy of the country in whose market the monopoly operates.

- We will examine two standard regulatory methods:

a) Rate of Return Regulation (ROR)

Fendekova and Fendek [6] show, that this method is used to regulate the output prices of power and gas companies in most developed economies. The essence is to set a price of production or service so that the regulated entity pays all its reasonable and prudent costs from its sales as well as a regulated return on its prudent investment. If the previous test year is used as the basis for the model database, the regulatory authority shall base its data on the last financial year. The identified costs for the supply and sale of energy are subsequently corrected by "known and measurable changes" (eg, input price increases) reflecting changes in costs and sales between the end of the trial year and the year in which tariffs will begin to pay.

b) Performance Based Regulation (PBR).

PBR is actually a modified ROR control. Under this regime, the regulator sets out the basic revenue requirements and basic tariffs when the PBR is implemented for the first time. At a predetermined interval (most often from 3 to 10 years), yield requirements and basic tariffs are reviewed on ROR principles. Tariff revisions are adjusted periodically according to specific schemes. The PBR mode removes some regulatory interventions and checks on ROR regulation and replaces their motivation and sanctioning system with the performance of the regulated entity by departing from the "normal" predicted state.

In Shy [10] we can see, that typical methodological tool for price regulation applied by price regulators to set maximum price of network industries products is the regulation on the basis of the rate of return – Rate of Return Regulation, by which are the prices of electricity, gas and other companies regulated in the most of the developed countries.

The aim is to ensure that the regulated entity will set the price of commodity or service for its consumers in a way the revenues will cover all its *reasonable and provident costs incurred* as well as regulated return on its *provident*[3] investment.
Let us now analytically derive the allowable rate of cost return for investment, co called *RoR* parameter of the regulated entity. Let us suppose that the firm is producing a homogenous product in production volume *Q*, which it realizes on a relevant market for the price *P*. Let us further suppose that the firm uses two production factors, namely labor force with consumption level *L* by labor price *w* and the capital with consumption level *K* by the capital price *r*.

The profit of the firm is generally defined as the difference between the yields and costs

$$\pi(Q) = R(Q) - C(Q)$$

where

R(Q) = $P \times Q$ – function of revenues of the firm, *R(Q)*: $R \to R$
C(Q) = $w \times L + r \times K$ – function of the total costs of the firm, *C(Q)*: $R \to R$

If we substitute general cost function on the basis of consumption of production factors, we get a profit function in a following form

$$\pi(q) = p \times q - w \times L - r \times K$$

If we further express the production volume *Q* on the basis of the production function in the form

$$Q = f(K, L)$$

and the production price *P* on the basis of the price-demand function in the form

$$P = p(Q)$$

then we can express the profit function in the form

$$\pi(Q) = p(f(K, L)) \times f(K, L) - w \times L - r \times K$$

A non-regulated firm can set its endogenous decision parameters in any way. So it chooses an optimum output volume Q^*, an acceptable optimum price P^* and corresponding consumption levels of the production factors labor *L* and capital *K* in a way to reach maximum profit. Optimum output and optimum price will be calculated by solving the following mathematical programming task

$$\pi(Q) = p(f(K, L)) \times f(K, L) - w \times L - r \times K \to \max$$

$$K, L \in R_{\geq 0}$$

In this case the non-regulated firm has no formal boundaries for setting the parameters guaranteeing its maximum profit. On the other hand, the regulated firm must respect boundaries given by the regulator. Price regulation regime on the basis of the rate of return lies in a fact, that through the use of exogenously defined control variable *RoR* the allowable level of quotient of the revenues $w \times L$ reduced by its non-capital expenditures $L \times w$ and the volume of consumed capital *K* is regulated.

In other words, a firm can optimize or freely determine the consumption levels of labor *L*, capital *K* by the market prices of production factors *w*, *r* and on the other side the level of its production *Q* but also the production price *P*. However firm has to respect the rate of return defined by the regulator i.e. the validity of the relation

$$RoR \geq \frac{P \times Q - w \times L}{K}$$

Let us further in detail explore the relation between the rate of return of the capital expenditures and the profit of the regulated entity. Profit can be analytically expressed as the difference between the proceeds and the costs of the firm in the form

$$\pi(Q) = P \times Q - w \times L - r \times K \qquad (1)$$

[3] It should be noted, that under term *provident* we understand situations, when certain decision about investment or expenditures has been made in conditions of verified and relevant information available in the time of decision making.

Let us deduct the price of the capital r from the both sides of the relation (1). We get the relation

$$RoR - r \geq \frac{P \times Q - w \times L}{K} - r$$

After modification we get

$$RoR - r \geq \frac{P \times Q - w \times L - r \times K}{K} \tag{2}$$

From the comparison of the relation (2) and (3) we get the relation

$$(RoR - r) \times K \geq \pi(Q) \tag{3}$$

From the relation (3) we can see that the regulated entity can set its system parameters only in a way that profit it reaches will not exceed the value of the capital evaluated by the difference between the rate of return *RoR* defined by the regulator and the price of the capital *r*.

Regulated entity can set its controlled (endogenous) decision parameters only in a way respecting regulatory condition. It sets the regulated volume of output Q_R, the acceptable regulated price P_R and the corresponding consumptions of production factors labor L and capital K in a way to reach maximum profit while respecting regulatory conditions (3) about non-exceeding the reasonable level of profit. The regulated output and regulated price are calculated by solving of the following mathematical programming task

$$\pi(q) = p(f(K,L)) \times f(K,L) - w \times L - r \times K \rightarrow \max$$

subject to (4)

$$p(f(K,L)) \times f(K,L) - w \times L - r \times K - (RoR - r) \times K \leq 0$$

$$K, L \in R_{\geq 0}$$

Solution of this optimalization task is optimum consumption level of production factors labor L^* and capital K^*. On their basis with the help of production function the regulated optimum level of output Q_R^* is quantified according to the following form $Q_R^* = f(K^*, L^*)$ and regulated optimum price P_R^* with the help of price-demand and production function on the basis of the relation $P_R^* = p(q_R^*) = p(f(K^*, L^*))$, while respecting the rate of return on capital defined by the *RoR* parameter, i.e. exogenous parameter set by the regulator.

We can show, that firm can produce regulated production volume Q^*_R with any combination of variable inputs labor and capital corresponding to isoquant $Q^*_R = f(L, K)$. So if the regulation on the basis of the rate of return stimulates firm towards using large volumes of capital in order to reach maximum profit allowed, the firm is not motivated to use often more effective and from the point of view of production price and production volume of inputs equivalent combinations of inputs localized on isoquant (e.g. supporting employment) in comparison with purposeless investments into the equipment. However there are also other, more sophisticated examples of cost structure of the firm.

This scheme of regulated firm's decision making described and theoretically justified by H. Averch a L. Johnson is known in literature as Averch – Johnson effect. Term Averch – Johnson effect is used to identify the inclination of the firm regulated on the *RoR* basis to use inappropriate ration of capital to labor to create certain production volume. Thorough analyses were elaborated on the basis of this principle.

Fendekova and Fendek [6] show, that on the basis of the Averch-Johnson model analysis several important conclusions about the rate of return regulated firms' behavior are elaborated:

1. Rate of return on capital regulated firm in effort to raise its allowed reasonable profit is motivated to inappropriate and wasteful increase of capital investments.

2. When the rate of return on capital investments is decreased, assuming $RoR > r$ is still valid, firm in order to maintain profit volume rises capital expenditures.

3. We can tell that decreasing the rate of return on capital investment means tightening the regulation conditions for regulated firm.

4 Conclusion

In The aim of this paper was to analyze the effects of network industries price regulation in Slovak Republic from the point of achieved level of reasonable profit of the regulated entities.

On the basis of the firms behavior in the conditions of the rate of return on capital regulation we show that in this regulatory scheme the firm has the tendency to respond to the tightening of the regulatory conditions by increasing the volume of used capital. Increase of used capital volume is, however, not the goal that the regulatory system is following. Regulation aims to influence others, for the firm and for the economy more important indicators, such as the volume of production, level of sales or the volume of costs.

Therefore other price regulation forms were developed to influence regulated firm reasonable profit on the basis of its total costs. The aim is, with the help of the regulatory mechanisms, to encourage effective development of the regulated entity.

On the basis of the Slovak Republic economic development analysis, we can state that the applied regulatory methods have not only the stabilizing effect in the network industries market environment but also the motivational aspect for the effective development of the regulated entities.

Acknowledgements

Supported by the grant No. 1/0697/15 *Optimization models and methods as tools of effective regulation in the modern theory of network industries organization* of the Grant Agency VEGA, Slovakia. (Scientific Grant Agency of the Ministry of Education, science, research and sport of the Slovak Republic and the Slovak Academy of Sciences).

References

[1] Bodea, T. and Ferguson, M.: Segmentation, Revenue Management, and Pricing Analytics. New York and London: Routledge 2015.
[2] Blum, U., Müller, S., and Weiske, A.: Angewandte Industrieökonomik: Theorien − Modelle – Anwendngen. Verlag Gabler, Berlin, 2006.
[3] Carlton, D.W. and Perloff, J. M.: Modern Industrial Organization. Boston: Addison Wesley, 2005.
[4] Fendek, M., and Fendeková, E.: Kuhn –Tucker optimality conditions in equlibrium models of network industry markets. *Politická ekonomie: teorie, modelování, aplikace.* Vysoká škola ekonomická, Praha **60**, 6, (2012), 801−821.
[5] Fendek, M., and Fendeková, E.: Microeconomic analysis of cartel equilibrium model and the examples of cartel agreements solutions at the Antimonopoly office of the Slovak Republic. In The Global Business & Economics Anthology: a publication of the B&ESI. - Danvers, MA, USA: Business & Economics Society International, (2012), 364-375.
[6] Fendek, M. and Fendeková, E.: Mikroekonomická analýza. IURA Edition. Bratislava. 557 strán. ISBN 80-88715-54-7, (2008).
[7] Meehen, J. M. et al.: Pricing and Profitability Management. Singapore: John Wiley & Sons, 2011.
[8] Nagle, T. T., Hogan, J. E., and Zale, J.: The Stratgy and Tacics of Pricing. Boston: Prentice Hall, 2011.
[9] Pepall, L., Richards, D., and Norman, D.: Industrial Organization: Contemporary Theory and Empirical Applications.New York: Wiley-Blackwell, 2008.
[10] Shy Oz: How to Price. A Guide to Pricing Technics and Yeld Management. New York: Cambridge Univesity Press, 2008.

Game theory models of co-opetition

Petr Fiala[1]

Abstract. Co-opetition is a business strategy that goes beyond the rules of competition and cooperation to combine the advantages of both. Co-opetition combines the advantages of both competition and cooperation into new dynamic, which can be used to not only generate more profits but also to change nature of the business environment in benefit of users. Traditional game theory is divided into non-cooperative and cooperative models. Biform games combine non-cooperative and cooperative models. The goal of coopetition is to move the players from a zero-sum game to a plus-sum game, a scenario in which the end result is more profitable when the competitors work together. The paper presents game theory models that represent the co-opetition concept. The coopetition business model has PARTS of a business strategy - five dimensions a company can use to identify strategies that change the game: Players, Added value, Rules, Tactics, and Scope. The players are the firm, customers, suppliers, competitors and complementors, competitors whose products add value. An important part of the game is to learn which conditions will influence the players to either compete or cooperate.

Keywords: co-opetition, game theory, dynamics, multiple criteria

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The traditional concept of business as a "winner takes all" contest is replaced by a new concept. Co-opetition is a business strategy that goes beyond the rules of competition and cooperation to combine the advantages of both. Co-opetition combines the advantages of both competition and cooperation into new dynamic, which can be used to not only generate more profits but also to change nature of the business environment in benefit of users (see [4]).

The co-opetition is based on game theory. The work of John von Neumann and Oskar Morgenstern [20] is the classic work upon which modern game theory is based. Since then, an extensive literature on game theory was published. For example, Myerson's book [16] provides a clear and thorough examination of the models, solution concepts, results, and methodological principles of non-cooperative and cooperative game theory. Game theory models analyze situations where players make decisions to maximize their own utility, while taking into account that other players are doing the same, and that decisions, made by players, impact others utilities. Traditional game theory is divided into non-cooperative and cooperative models. The non-cooperative theory of games is strategy oriented; it studies what one may expect the players to do. The non-cooperative theory is a "micro" approach in that it focuses on precise descriptions of what happens. The co-opetition concept combines non-cooperative and cooperative game theory models.

Dynamics is a very important part of co-opetition models. Current business conditions are changing rapidly. New products are evolving faster. Searching for relationships with complementors brings ever new opportunities that bring added values. This is one of the basics of coopetition. This dynamics must be included in the new cooperative models.

Searching for new opportunities with complementors and negotiations with competitors are carried out in the face of multiple criteria. There are multiple criteria in co-opetition models, such as economic, social, environmental, technological, and others. Multicriteria evaluation approaches will be tested for inclusion in the model of co-opetition.

The rest of the paper is organized as follows. Section 2 presents basics of co-opetition model. Section 3 summarizes the basics of the game theory concepts. Biform games as basic models for modeling of co-opetition problems are analyzed in Section 4. Extensions of biform games are presented in Section 5. Section 6 presents conclusions.

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech Republic, e-mail: pfiala@vse.cz

2 Co-opetition business model

The co-opetition business model has PARTS of a business strategy - five dimensions a company can use to identify strategies that change the game:

- Players,
- Added value,
- Rules,
- Tactics,
- Scope.

The players are the firm, customers, suppliers, competitors and complementors (competitors whose products add value). An important part of the game is to learn which variables will influence the players to either compete or cooperate. Added values are given through complementors. Rules structure negotiations between buyers and sellers. Tactics are actions taken to shape other players' perceptions. Scope means recognizing the links between games through Players, Added values, Rules, and Tactics.

The concept of co-opetition has been refined and applied in many cases. The paper [21] predicts that strategic planners in organizations of the future need to consider the potential benefits of collaborating, co-operating and co-ordinating with others serving the same markets, rather than pursuing conventional "competition". The paper [10] examines empirically the impact of co-opetition on firm competitive behavior. The co-opetition concept with technological innovation is designed for small and medium-sized enterprises (see [9]) and for large firms (see [11]). Specific issues of co-opetition were analyzed, such as dynamics (see [2]), tension between cooperation and competition (see [3]), and information sharing (see [18]). The co-opetition is based on game theory (see [17]). The players are connected to the supply chain. The paper [15] describes coordination game model of co-opetition relationship on cluster supply chains. The paper [12] analyzes impact of product pricing and timing of investment decisions on supply chain co-opetition. There are some critical reviews of the co-opetition concept (see [1]). Deeper use of game theory, the theory of supply chain and other disciplines is required for new more precise models of co-opetition.

3 Basic game theory concepts

We start from the basic non-cooperative and cooperative concepts of the game theory that are applied in the proposed approach for co-opetition models (see [6]).

Most non-cooperative allocation strategies in distributed systems consist of following steps:

- The formulation of utility (pay-off) functions for the system participants.
- The formulation of best response strategies.
- The existence of Nash equilibrium is proved in the system of multiple agents.
- Efficiency is measured compared to achievable welfare.

We start from the basic non-cooperative and cooperative concepts of the game theory that are applied in the proposed approach for co-opetition models.

An *n*-player non-cooperative game in the normal form is a collection

$$\{N = \{1,2,\dots,n\};\ X_1, X_2, \dots, X_n;\ \pi_1(x_1, x_2, \dots, x_n), \pi_2(x_1, x_2, \dots, x_n), \dots, \pi_n(x_1, x_2, \dots, x_n)\}, \quad (1)$$

where N is a set of n players; X_i, $i = 1, 2, \dots, n$, is a set of strategies for player i; $\pi_i(x_1, x_2, \dots, x_n)$, $i = 1, 2, \dots, n$, is a pay-off function for player i, defined on a Cartesian product of n sets X_i, $i = 1, 2, \dots, n$.

Decisions of other players than player i are summarized by a vector

$$\mathbf{x}_{-i} = (x_1, \dots, x_{i-1}, x_{i+1} \dots, x_n). \quad (2)$$

A vector of decisions $(x_1^0, x_2^0, \dots, x_n^0)$ is Nash equilibrium of the game if

$$x_i^0(\mathbf{x}_{-i}^0) = \underset{x_i}{\operatorname{argmax}}\, \pi_i(x_i, \mathbf{x}_{-i}) \forall i = 1,2, \dots, n. \quad (3)$$

Nash equilibrium is a set of decisions from which no player can improve the value of his pay-off function by unilaterally deviating from it.

Cooperative game theory looks at the set of possible outcomes, studies what the players can achieve, what coalitions will form, how the coalitions that do form divide the outcome, and whether the outcomes are stable and robust.

The maximal combined output is achieved by solving the following task

$$\mathbf{x}^0 = \underset{\mathbf{x}}{\operatorname{argmax}} \sum_{i=1}^{n} \pi_i(x_i). \tag{4}$$

Allocation mechanisms are based on different approaches such as negotiations, auctions, Shapley values, etc. When modeling cooperative games is advantageous to switch from the game in normal form to the game in the characteristic function form. The characteristic function of the game with a set N of n players is such function $v(S)$ that is defined for all subsets $S \subseteq N$ (i.e. for all coalition) and assigns a value $v(S)$ with following characteristics:

$$v(\varnothing) = 0, \quad v(S_1 \cup S_2) \geq v(S_1) + v(S_2), \tag{5}$$

where S_1, S_2 are disjoint subsets of the set N. The pair (N, v) is called a cooperative game of n players in the characteristic function form.

A particular allocation policy, introduced by Shapley (see [19]) has been shown to possess the best properties in terms of balance and fairness. So called Shapley vector is defined as

$$\mathbf{h} = (h_1, h_2, ..., h_n), \tag{6}$$

where the individual components (Shapley values) indicate the mean marginal contribution of i-th player to all coalitions, which may be a member. Player contribution to the coalition S is calculated by the formula

$$v(S) - v(S - \{i\}). \tag{7}$$

Shapley value for the i-th player is calculated as a weighted sum of marginal contributions according to the formula:

$$h_i = \sum_{S} \left\{ \frac{(|S|-1)!\,(n-|S|)!}{n!} . \left[v(S) - v(S - \{i\}) \right] \right\}, \tag{8}$$

where the number of coalition members is marked by symbol $|S|$ and the summation runs over all coalition $i \in S$.

4 Biform games

We will use biform games as basic models for modeling of co-opetition problems. A biform game is a combination of non-cooperative and cooperative games (see [5]). The goal of co-opetition is to move the players from a zero-sum game to a plus-sum game, a scenario in which the end result is more profitable when the competitors work together. We propose to divide biform games into sequential and simultaneous.

Sequential biform games

The sequential biform game is a two-stage game: in the first stage, players choose their strategies in a non-cooperative way, thus forming the second stage of the game, in which the players cooperate. First, suppliers make initial proposals and take decisions. This stage is analyzed using a non-cooperative game theory approach. The players search for Nash equilibrium by solving the problem (3).

Then, players negotiate among themselves. In this stage, a cooperative game theory is applied to characterize the outcome of negotiation among the players over how to distribute the total surplus. Each player's share of the total surplus is the product of its added value and its relative negotiation power. We propose to distribute the total surplus to players by Shapley values (8).

The sequential biform game is used f. e. in supply chains (see [7] and [8]). In the non-cooperative part, a coordination mechanism based on a specific buy-back contract is applied between producers and customers with price-dependent stochastic demand. The contract has desirable features: full coordination of the supply chain, flexibility to allow any division of the supply chain's profit, and easy to use. The cooperative part is merely focused on two concepts, coalition formations by resource capacity constraints and profit sharing. Profit sharing is carried out on the recognized concept of Shapley value.

Simultaneous biform games

The simultaneous biform game is a one-stage model where combinations of concepts for cooperative and non-cooperative games are applied. The combinations will be changed according situations in co-opetition problems. The first problem is a classification of situations. The situations are affected by:

- which players can cooperate,
- to what scope they can cooperate.

If all players can cooperate fully, then a standard cooperative model can be used (4) with subsequent distribution of the result according to the Shapley values (8). If no one can cooperate even in a partial content, a standard non-cooperative model is used (3). These are two extreme cases.

The players in the co-opetition model are the firm, customers, suppliers, competitors and complementors (competitors whose products add value). The relationship between the firm and the direct competitors is non-cooperative. The relationship between the firm and the complementors in a search for common added values is cooperative. The relationship between the firm and the suppliers can be partly cooperative; for some criteria cooperative (f. e. timing of deliveries), for others non-cooperative (f. e. price).

The scope of cooperation is determined by the various constraints that result from the fact that players are under internal and external pressures. The effects of pressures will be reflected in restrictive conditions. The scope of cooperation is dynamic and given by multiple criteria.

5 Extensions of biform games

We propose new models for co-opetition with players, added values, rules, tactics, and scope. Players are negotiation participants with their tactics when searching for added values. Participants negotiate under pressure by given rules with changing scope.

The traditional game concepts assume a fixed structure and fixed sets of strategies. The co-opetition concept expects changes due to the expansion of opportunities for cooperation with complementors and the changes over time in general. Sets of strategies will be taken as dynamic $X_i(t)$, for players $i = 1, 2, \ldots, n$, depending on the discrete time periods $t = 1, 2, \ldots, T$. Dynamic evaluations of strategies will be also considered.

There are multiple criteria in co-opetition models, such as economic, environmental, technological, and others. Multicriteria evaluation approaches are tested for inclusion in the model of co-opetition. The problems with multiple criteria will be solved by specific multi-criteria decision support systems (e.g. IZAR – see [14]) or by spreadsheets (see [13]).

Negotiation models

We summarize basic notations of negotiation models and then describe pressure concept of negotiation. Suppose n negotiation participants. Denote X as decision space for the negotiating process. Elements of this space are decisions $\mathbf{x} \in X$, which are vectors whose components represent the parameters of the decision. A consensus decision $\mathbf{x}^*$ should be chosen from the decision space X. Each participant evaluates decisions by several criteria and compares the decisions with the target values. The criteria are in the form of criteria functions, that all participants want to maximize their values. Each participant in negotiations may have a different number of criteria. Denote $\mathbf{f}^1(\mathbf{x})$, $\mathbf{f}^2(\mathbf{x})$, ..., $\mathbf{f}^n(\mathbf{x})$ vector criteria functions that transform decision $\mathbf{x}$ into the vectors of target values $\mathbf{y}^1$, $\mathbf{y}^2$, ..., $\mathbf{y}^n$ of the target spaces of the participants Y^1, Y^2, ..., Y^n. These achievements, however, the participant tries to not reveal his interests and his strategy to all players. Own negotiations and exchanges of information between participants are happening in the decision space.

The negotiation process can be represented by dynamic models. Individual time moments correspond to states of negotiation, in which the current joint problem representation shows the degree of consensus or conflict between the parties to the negotiations. Developments of problem representations can be described as a search for consensus through the exchange of information between participants. The negotiation process is dynamic, and suppose that there is at discrete time points $t = 1, 2, ..., T$. At the time T the process is completed by finding a trajectory to time horizon T. Negotiation process over time can be modeled as a gradual change of the negotiation space, which is a subset of the decision space containing acceptable decisions of participants in the negotiation time until a single element negotiation space is reached.

For each participant we can formulate a set of acceptable decisions, which is a set of decisions that are permissible and acceptable in terms of the required aspiration levels of criteria functions. The aspiration levels $\mathbf{b}^i(t)$, $i = 1, 2, ..., n$, $t = 1, 2, ..., T$, of criteria functions represent opportunities for added values. At the beginning of the negotiations it has the form

$$X_i(0) = \{\mathbf{x};\ \mathbf{x} \in X,\ \mathbf{f}^i(\mathbf{x}) \leq \mathbf{b}^i(0)\},\ i = 1, 2, ..., n. \tag{9}$$

Then we can define the negotiation space at the beginning of the negotiations as an intersection of sets of acceptable decisions of all participants in negotiations

$$X_0(0) = \bigcap_{i=1}^{r} X_i(0) \tag{10}$$

If the negotiation space $X_0(0)$ is a single element set, then the negotiation problem is trivial. This element is the consensus. Negotiation problem becomes interesting when the negotiation space is empty or contains more than one element. In the first case, participants have to reduce some or all of the aspiration levels of criteria functions, but participants are involved in the reduction of certain criteria more and other less. In the latter case, each element of the negotiation space is acceptable to all participants, but different elements are evaluated differently, because they meet the criteria of the participants on different levels. Further negotiations are conducted in time points $t = 1, 2, ..., T$, and should lead to a consensus decision, to achieve single-element negotiation space $X_0(t)$.

Concept of pressure

This pressure negotiation concept is based on the assumption that each participant decides under pressure of objective context, subject to a variety of internal and external pressures. Participant is under pressure, for example, if he wants to reach a consensus, he is aware of prices for delayed decisions; other participants influence their behavior, etc. Pressure is a term that includes internal values and external influences and determines the decision making process. Assume that the pressure does not affect the selection of decisions directly, but through a set of conditions that have to be satisfied. Then we can consider the effects of pressure, which is reflected in changes in the set of constraints. This leads to a change of set of acceptable participants' decisions and a change of the negotiation space and can lead to a consensus.

Pressures acting on the aspiration levels of criteria functions that change in time points $t = 1, 2, ..., T$, and thus change the set of acceptable decisions

$$X_i(t) = \{\mathbf{x};\ \mathbf{x} \in X,\ \mathbf{f}^i(\mathbf{x}) \leq \mathbf{b}^i(t)\},\ i = 1, 2, ..., n. \tag{11}$$

Changes of aspiration levels are described by a vector of values $\mathbf{p}^i(t)$ at time t ($\mathbf{p}^i(0) = \mathbf{0}$)

$$\mathbf{b}^i(t) = \mathbf{b}^i(t\text{-}1) + \mathbf{p}^i(t). \tag{12}$$

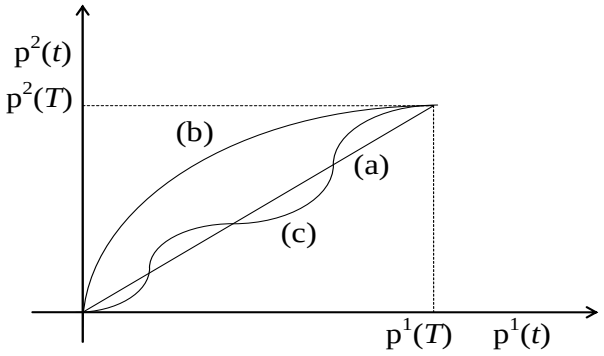


Figure 1: Trajectory of pressures

Vector $\mathbf{p}^i(t)$ describes the changes in aspiration levels of the i-th participant at time t. Vector $\mathbf{p}(t)$ describes the changes of all aspiration levels for all participants at time t, this vector has as many components as there are together all soft constraints. By joining points $\mathbf{p}(0)$, $\mathbf{p}(1)$, ..., $\mathbf{p}(T)$ yields a continuous vector function $\mathbf{p}(t)$ defined on the interval $< 0, T >$, which is called trajectory of pressures and represents tactics.

There are several different trajectories of pressures that lead to the same consensus decision (see Figure 1). Trajectory (a) corresponds to negotiation with constant pressure effects. Trajectory (b) represents the negotiation with the effects of pressure nonlinearly dependent on time. If the parties negotiate increases and reduces the aspiration levels, this type of negotiation corresponds to trajectory (c).

6 Conclusions

New models of co-opetition are developed. Relevance of the project can be seen at two levels: theoretical and practical. The theoretical contribution can be seen in the creation of new models for biform games that will capture the sequential or the simultaneous use of non-cooperative and cooperative models. The models will include dynamic changes of parameters in the games, unlike traditional models of game theory. Multiple criteria evaluation of strategies can be also included in the models.

The practical contribution consists in the use of models for applications. Co-opetition business strategy is considered to be very promising. Changes in the new models will better reflect the real conditions. The proposed

models can be used not only for co-opetition strategy, but also for many other decision problems which combine the advantages of competition and cooperation.

Acknowledgements

Supported by the grant No. P402/12/G097 (DYME – *Dynamic Models in Economics* of the Grant Agency of the Czech Republic and by the grant No. IGA F4/57/2017, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Armstrong J. S.: Review of Barry J. Nalebuff and Adam N. Brandenburger, Co-opetition 1. Revolutionary Mindset that Redefines Competition and Co-operation 2. The Game Theory Strategy that's Changing the Game of Business. *Journal of Marketing* **61** (1997), 92-95.

[2] Bengtsson, M., Eriksson, J., Wincent, J.: Co-opetition dynamics–an outline for further inquiry. *Competitiveness review: An international business journal* **20** (2010), 194-214.

[3] Bengtsson, M., & Kock, S.: Tension in Co-opetition. In *Creating and Delivering Value in Marketing*. Springer International Publishing, Berlin, 2015, 38-42.

[4] Brandenburger, A. M., Nalebuff, B. J.: *Co-opetition*. Crown Business, New York, 2011.

[5] Brandenburger, A., Stuart, H.: Biform games. *Management science* **53** (2007), 537-549.

[6] Dlouhý, M., Fiala, P.: *Teorie ekonomických a politických her*. Oeconomica, Praha, 2015.

[7] Fiala, P.: Biform games in supply chains. In: *Mathematical Methods in Economics* 2015. University of West Bohemia, Pilsen, 177–183.

[8] Fiala, P.: Profit allocation games in supply chains. *Central European Journal of Operations Research* **24** (2016), 267-281.

[9] Gnyawali, D. R., Park, B. J. R.: Co-opetition and technological innovation in small and medium-sized enterprises: A multilevel conceptual model. *Journal of small business management* **47**(2009), 308-330.

[10] Gnyawali, D. R., He, J., & Madhavan, R.: Impact of co-opetition on firm competitive behavior: An empirical examination. *Journal of Management* **32**(2006), 507-530.

[11] Gnyawali, D. R., & Park, B. J. R.: Co-opetition between giants: Collaboration with competitors for technological innovation. *Research Policy* **40** (2011), 650-663.

[12] Gurnani, H., Erkoc, M., & Luo, Y. Impact of product pricing and timing of investment decisions on supply chain co-opetition. *European Journal of Operational Research*, **180** (2007), 228-248.

[13] Jablonsky, J.: MS Excel based software support tools for decision problems with multiple criteria. In: 17th *International Conference on Enterprise and the Competitive Environment* (ECE) *Location*. Mendel Univ, Fac Business & Econ, Brno, 2014, 251-258.

[14] Kalčevová, J., Fiala, P.: IZAR – Multicriteria Decision Support System. In: *Mathematical Methods in Economics* 2006. University of West Bohemia, Pilsen, 277–282.

[15] Min, Z., Feiqi, D., Sai, W.: Coordination game model of co-opetition relationship on cluster supply chains. *Journal of systems engineering and electronics* **19** (2008), 499-506.

[16] Myerson, R. B.: Game Theory: Analysis of Conflict. Harvard University Press, Cambridge, 1997.

[17] Okura, M., Carfì, D.: Coopetition and Game Theory. *Journal of Applied Economic Sciences* **9** (2014), 1-29.

[18] Renna, P., & Argoneto, P.: Capacity investment decision in co-opetitive network by information sharing. *Computers & Industrial Engineering*, **62** (2012), 359-367.

[19] Shapley, L. S.: A value for n-person games. In: *Contributions to the Theory of Games* II. (Tucker, A. W., Luce, R. D., eds.). Princeton University Press, Princeton, 1953, 307-317.

[20] von Neumann, J., Morgenstern, O.: Theory of games and economic behavior. Princeton University Press, Princeton, 1944.

[21] Zineldin, M.: Co-opetition: the organisation of the future. *Marketing Intelligence & Planning* **22**(2004), 780-790.

Implicit-explicit scheme combined with wavelets for pricing European options

Václav Finěk [1]

Abstract. Option products are frequently traded in the financial markets. To price these options, advanced mathematical models are employed, yielding multidimensional parabolic partial differential equations of the convection-diffusion type. And it is necessary to develop efficient, stable, and robust numerical methods to solve them. We start here with the one-dimensional Black-Scholes equation.
We employ the implicit-explicit scheme for time discretization and wavelets for space discretization. First, we split the standard weak form of the Black-Scholes operator into an symmetric part and into an unsymmetric part. And then we treat implicitly the symmetric part of the weak form of the Black-Scholes operator and explicitly its unsymmetric part. Consequently, the arising system of equations can be efficiently preconditioned using the standard wavelet based preconditioning. Numerical examples are given.

Keywords: Black-Scholes equation, European option, wavelets, implicit-explicit time discretization.

JEL classification: C63, G13
AMS classification: 35K20, 65T60

1 Introduction

Options are frequently traded in the financial markets. To price these options, advanced mathematical models are employed, yielding multidimensional parabolic partial differential equations of the convection-diffusion type. And it is necessary to have at one's disposal efficient, stable, and robust numerical methods to compute their fair values. This paper is concerned with numerical solution of the one-dimensional Black-Scholes equation. We apply the implicit-explicit scheme for time discretization and Hermite cubic spline wavelets with four vanishing moments for space discretization. The proposed implicit-explicit scheme is based on the operator splitting method. We split the arising system into an symmetric part and into an unsymmetric part. And then we treat implicitly the symmetric part of the weak form of the Black-Scholes operator and explicitly its unsymmetric part. Consequently, the arising system of equations can be efficiently preconditioned using the standard wavelet based preconditioning [6] and solved by the conjugate gradient method with small number of iterations. Numerical examples are given.

For an overview of classical methods for numerical solution of the Black-Scholes equation, we refer to [1]. Recently also new approaches such as the discontinuous Galerkin method [11, 12] or wavelets [4, 5] have been used for an efficient solution of different models. In [7], we proposed a construction of the wavelet basis with respect to which the Black-Scholes operator is sparse even in the case when volatilities are piecewise quadratic polynomials. In [9], we used the above described operator splitting method with $\theta-$scheme and we showed that arising system of equations can be efficiently preconditioned using an wavelet based preconditioning and consequently efficiently solved by the conjugate gradient method (CGS). At each time step and at each decomposition level, it was necessary to perform only few iterations of CGS to have the norm of the residuum smaller than 10^{-8}. The advantages of this approach consist in higher order accuracy, a small number of iterations needed to resolve the problem with desired accuracy and a high potential in adaptive methods due to the four vanishing wavelet moments.

As reported for example in [10] even for European call options in one space dimension, the agreement between the numerical solution and the analytical solution can be quite good, but this has not to be the case for the first and the second order derivatives. They used Crank-Nicholson method for time discretization and finite differences for space discretization. Even worse was the situation for the European digital options with no convergence for option values in the maximum norm. To show that this behavior is representative also for more realistic applications, they also employed Heston stochastic volatility model or European call options in two space dimensions. This behavior is due to the initial data irregularities in the model and therefore Crank-Nicolson scheme is often used together with Rannacher startup procedure to improve convergence results. For more details, we refer to [10]. Therefore it is still important to investigate numerical methods for the one-dimensional Black-Scholes equation.

[1]Department of Mathematics and Didactics of Mathematics, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic, Vaclav.Finek@tul.cz

2 Black-Scholes equation

Let $x > 0$ be a price of the underlying assets, t represent time to maturity, T be an expiration time, r be a risk-free interest rate, and σ be an implied volatility. Then the market price $u(x,t)$ of the option at time t with a price x of the underlying assets can be computed as the solution of the Black-Scholes equation [2]:

$$\frac{\partial u}{\partial t} + \mathcal{L}_{BS}(u) = 0, \quad \text{in } (0, S_{\max}) \times (0,T), \tag{1}$$

where $S_{\max}$ is a large enough constant coming from the approximation of the unbounded interval $(0,\infty)$ and the Black-Scholes operator $\mathcal{L}_{BS}$ is given by

$$\mathcal{L}_{BS}(u) = -\frac{\sigma^2 x^2}{2}\frac{\partial^2 u}{\partial x^2} - rx\frac{\partial u}{\partial x} + ru. \tag{2}$$

Boundary conditions are prescribed in this way:

$$u(0,t) = Ke^{-rt} \quad \text{and} \quad u(S_{\max},t) = 0 \qquad \forall t \in (0,T), \tag{3}$$

where K is a strike price. We consider here only put options. Then the initial condition is given by a piecewise linear payoff function:

$$u(x,0) = \max(K - x, 0). \tag{4}$$

Further, we transform the problem (1) with non-homogeneous Dirichlet boundary conditions to the problem

$$\frac{\partial v}{\partial t} + \mathcal{L}_{BS}(v) = f, \quad \text{in } (0, S_{\max}) \times (0,T), \tag{5}$$

with homogeneous Dirichlet boundary conditions, where

$$u = v + w \quad \text{and} \quad f(w) = -\frac{\partial w}{\partial t} - \mathcal{L}_{BS}(w), \tag{6}$$

with

$$w(0,t) = Ke^{-rt} \quad \text{and} \quad w(S_{\max},t) = 0 \qquad \forall t \in (0,T). \tag{7}$$

To solve the problem (5), we apply the $\theta-$scheme for time discretization and wavelets for space discretization. The standard weak form of the Black-Scholes operator $\mathcal{L}_{BS}$ is the following:

$$(\mathcal{L}_{BS}(u), z) = \frac{\sigma^2}{2}\int_0^{S_{\max}} x^2 \frac{\partial u}{\partial x}\frac{\partial z}{\partial x}\,dx + \left(\sigma^2 - r\right)\int_0^{S_{\max}} x\frac{\partial u}{\partial x} z\,dx + r\int_0^{S_{\max}} uz\,dx, \tag{8}$$

where $z \in H_0^1(0, S_{\max})$ and $(\cdot,\cdot)$ denotes L^2 inner product. Now, we split this weak form into the symmetric part

$$a(u,z) := \frac{\sigma^2}{2}\int_0^{S_{\max}} x^2 \frac{\partial u}{\partial x}\frac{\partial z}{\partial x}\,dx + r\int_0^{S_{\max}} uz\,dx \tag{9}$$

and into the unsymmetric part

$$b(u,z) := \left(\sigma^2 - r\right)\int_0^{S_{\max}} x\frac{\partial u}{\partial x} z\,dx. \tag{10}$$

Then, we use the implicit-explicit scheme for time discretization with implicit treatment of the symmetric part and explicit treatment of the unsymmetric part. Let τ be a time step, $t_l = l\tau$, $l = 0, \ldots, \tau^{-1}T$, and further let us denote $v_l(x) = v(x,t_l)$, and $w_l(x) = w(x,t_l)$. The above operator splitting yields the following weak form:

$$\frac{(v_{l+1} - v_l, z)}{\tau} + \frac{1}{2}\left(a(v_{l+1},z) + a(v_l,z)\right) + b(v_l,z) = \frac{1}{2}\left(f(w_{l+1}) + f(w_l), z\right), \tag{11}$$

which leads after wavelet discretization in the space to well-conditioned symmetric systems of equations [6].

3 Hermite cubic spline wavelets

We use Hermite cubic spline wavelets with four vanishing moments for space discretization and we shortly review their basic properties. We consider families $\Psi = \{\psi_\lambda, \lambda \in \mathcal{J}\} \subset L_2(0,1)$ of wavelets such that $||\psi_\lambda||_{L_2} = 1$. Let $\mathcal{J}$ be an infinite index set and $\mathcal{J} = \mathcal{J}_\Phi \cup \mathcal{J}_\Psi$, where $\mathcal{J}_\Phi$ is a finite set representing scaling functions living on the coarsest scale. Any index $\lambda \in \mathcal{J}$ is of the form $\lambda = (j,k)$, where $|\lambda| = j$ denotes a scale and k denotes spatial location. Further, we will denote $\mathbf{D}^s$ a diagonal matrix, whose diagonal entries are $2^{s|\lambda|}$. Then $\mathbf{D}^{-s}\Psi = \{2^{-s|\lambda|}\psi_\lambda\}$ will denote a scaled wavelet basis. At last, for $s \geq 0$ the space H^s will denote a closed subspace of the Sobolev space $H^s(0,1)$, defined e.g. by imposing homogeneous boundary conditions at one or both endpoints, and for $s < 0$ the space H^s will denote the dual space $H^s := (H^{-s})'$. $\|.\|_{H^s}$ will denote the corresponding norm. Further $l_2(\mathcal{J})$ will denote the space consisting of the power summable sequences and $\|.\|_{l_2(\mathcal{J})}$ will denote the corresponding norm.

A family $\Psi = \{\psi_\lambda, \lambda \in \mathcal{J}\} \subset L_2(0,1)$ is called a *wavelet basis* of H^s for some $\gamma, \widetilde{\gamma} > 0$ and $s \in (-\widetilde{\gamma}, \gamma)$, if

- Ψ is a Riesz basis of H^s, that means Ψ forms a basis of H^s and there exist constants $c_s, C_s > 0$ such that for all $\mathbf{b} = \{b_\lambda\}_{\lambda\in\mathcal{J}} \in l_2(\mathcal{J})$ holds

$$c_s \|\mathbf{b}\|_{l_2(\mathcal{J})} \leq \left\|\mathbf{b}^T \mathbf{D}^{-s}\Psi\right\|_{H^s} \leq C_s \|\mathbf{b}\|_{l_2(\mathcal{J})}, \tag{12}$$

  where $\sup c_s$, $\inf C_s$ are Riesz bounds and $\operatorname{cond}\Psi := \dfrac{\inf C_s}{\sup c_s}$ is the condition number of Ψ.
- Functions are local in the sense that $\operatorname{diam}(\operatorname{supp}\psi_\lambda) \leq C2^{-|\lambda|}$ for all $\lambda \in \mathcal{J}$, where C is a constant independent of λ.
- Functions ψ_λ, $\lambda \in \mathcal{J}_\Psi$, have cancellation properties of order $m > 0$, i.e. $\forall v \in H^m(0,1)$:

$$\left|\int_0^1 v(x)\,\psi_\lambda(x)\,dx\right| \leq 2^{-m|\lambda|}\,|v|_{H^m(0,1)}.$$

  It means that integration against wavelets eliminates smooth parts of functions and it is equivalent with vanishing wavelet moments of order m.

As basis functions, Hermite cubic splines wavelets proposed in [8] are used because they have vanishing moments, short support, are well-conditioned and finally, for constant coefficient differential equations and also for the considered Black-Scholes operator, arising stiffness matrices are sparse in wavelet coordinates. In this special construction, four types of wavelets ψ_1, ψ_2, ψ_3, and ψ_4, are constructed and the wavelet basis is then formed by translations and dilations of these four types of wavelets. The first two types of wavelets have supports in $[-1,1]$ and are orthogonal to cubic polynomials. Moreover, the first one is odd and the second one is even. The second two types of wavelets have supports in $[-2,2]$. We impose again the orthogonality condition to cubic polynomials in intervals $[-2,0]$ and $[0,2]$ and one of them should be odd and the second one even. These properties ensure that both the mass and the stiffness matrix corresponding to the one-dimensional Laplacian as well as the stiffness matrices corresponding to the one-dimensional Black-Scholes equation have at most three wavelet blocks of nonzero elements in any column and then the number of nonzero elements in any column is bounded independent of matrix size. Condition numbers of these wavelets were later improved in [3]. As a consequence of properties of these wavelets, we obtain well-conditioned stiffness matrices with sparse structure and as well a sparse representation of the approximate solution. For more details, we refer to [3, 8].

4 Numerical results

In numerical experiments, we compute a price of a put option on the German DAX 30 stock market index. Based on the real data from 15 September 2011, we set the expiration date $T = 1/3$, the strike price 4000, implied volatility $\sigma = 0.4594$, and the risk-free interest rate $r = 0.0176$ [11]. The length of a computational domain depends on the maximal value of the German index DAX. We set here $S_{\max} = 16000$ to suppress the influence of the approximation of the unbounded interval $(0,\infty)$. In numerical experiments, we use the above mentioned wavelets. Let $\Psi = \{\psi_\lambda, \lambda \in \mathcal{J}\}$ be a wavelet basis of the space L_2 proposed in [8] and

$$\Psi_i = \{\psi_\lambda, |\lambda| \leq i, \lambda \in \mathcal{J}\}$$

be a wavelet basis containing all wavelets up to level i.

In the Table 1, i denotes the wavelet decomposition level, NTS denotes the number of time steps and order denotes computed order of convergence between two subsequent wavelet decomposition levels. In all numerical experiments, we used the conjugate gradient method to solve the arising system of linear equations. We set the first initial guess equal to the result obtained in the previous time level and we iterated until a relative residual was greater than $10^{-6}2^{-4i}$, where i is the wavelet decomposition level. Numerical results confirm that proposed scheme is the fourth order accurate in the space but we can observe slower convergence with respect to time.

Table 1 *Convergence history.*

i	NTS	rel. L_2-error	order	rel. H^1-error	order
1	4	1.07e-2		5.08e-2	
2	32	8.38e-4	3.67	7.40e-3	2.78
3	256	7.08e-5	3.57	1.30e-3	2.51
4	2048	6.48e-6	3.45	2.02e-4	2.69
5	16384	5.43e-7	3.58	2.74e-5	2.88

5 Conclusion

We applied the implicit-explicit scheme for time discretization and Hermite cubic spline wavelets with four vanishing moments for space discretization. The implicit-explicit scheme is based on the operator splitting method. We split the arising system into an symmetric part and into an unsymmetric part. And then we treat implicitly the symmetric part of the weak form of the Black-Scholes operator and explicitly its unsymmetric part. Consequently, the arising system of equations can be efficiently preconditioned using the standard wavelet based preconditioning and solved by the conjugate gradient method with small number of iterations. Numerical results confirm that proposed scheme is the fourth order accurate in the space but we can observe slower convergence with respect to the time.

Acknowledgements

This work was supported by grant No. GA16-09541S of the Czech Science Foundation.

References

[1] Achdou, Y., and Pironneau, O.: *Computational Methods for Option Pricing,* Society for Industrial and Applied Mathematics, Philadelphia, 2005.

[2] Black, F., and Scholes, M.: The pricing of options and corporate liabilities. *J. Polit. Econ.* **81**, pp. 637–659, 1973.

[3] Černá, D., and Finěk, V.: *On a sparse representation of a n-dimensional Laplacian in wavelet coordinates*, Result. Math. **69**, pp. 225-243, 2016.

[4] Černá, D.: *Numerical solution of the Black-Scholes equation using cubic spline wavelets,* in *AIP Conference Proceedings* 1789; Pasheva, V.; Popivanov, N.; Venkov, G. Eds.; American Institute of Physics, New York, 2016; No. 030001 (8 pages).

[5] Černá, D., and Finěk, V.: *Adaptive wavelet method for the Black-Scholes equation of European options,* in *34th International Conference Mathematical Methods in Economics (MME 2016)*; Kocourek, A; Vavrousek, M Eds.; Technical University of Liberec, Liberec, 2016; (6 pages).

[6] Černá, D., and Finěk, V.: *A diagonal preconditioner for singularly perturbed problems,* Bound. Value Probl. **2017:22**, DOI: 10.1186/s13661-017-0754-4, 2017.

[7] Černá, D., and Finěk, V.: *Sparse wavelet representation of differential operators with piecewise polynomial coefficients,* Axioms **6** (1), 4 (21pages), 2017.

[8] Dijkema, T. J., and Stevenson, R.: *A sparse Laplacian in tensor product wavelet coordinates*. Numer. Math. **115**, pp. 433–449, 2010.

[9] Finěk, V.: *Fractional step method for wavelet based solution of Black-Scholes equation,* in *AIP Conference Proceedings* 1789; Pasheva, V.; Popivanov, N.; Venkov, G. Eds.; American Institute of Physics, New York, 2016; No. 030007 (4 pages).

[10] Giles, M. B., and Carter, R.: *Convergence analysis of Crank-Nicolson and Rannacher time-marching*, J. Comput. Finance **9** (4), pp. 89-112, 2006.

[11] Hozman, J.: *Analysis of the discontinuous Galerkin method applied to the European option pricing problem*. in *AIP Conference Proceedings* 1570, Pasheva, V.; Popivanov, N.; Venkov, G. Eds.; American Institute of Physics, New York, 227 (2013); (8 pages).

[12] Hozman, J. and Tichý, T.: *DG method for numerical pricing of multi-asset Asian options - the case of options with floating strike*. Applications of Mathematics **62** (2), pp. 171-195, 2017.

Unconventional monetary policy in a small open economy under inflation targeting regime

Jakub Fodor[1], Osvald Vašíček[2]

Abstract. In this paper, we focus on consequences of unconventional monetary policy, namely, of interventions on foreign exchange (FX) market. In recent development of many economies, interest rate hit the zero lower bound, making traditional instrument of monetary policy unemployable. Depreciation of country's currency and subsequent interventions on FX market therefore provide for an option for inflation targeting in small open economies. We use a New-Keynesian dynamic stochastic model of general equilibrium. The possibility of central bank's interventions is modeled by adding a segment of FX dealers as a modified uncovered interest parity condition, according to [2]. The model is then estimated on data of Czech economy. We then introduce non-linearities to otherwise linear model, providing a way to model constraints on variables. Constraint on interest rate is chosen to model the position of the economy at zero lower bound, whereas constraint on exchange rate is used to model commitment of central bank to keep exchange rate above certain value. Model is solved using toolbox of [3], which introduces a way of dealing with non-linear models, by assuming perfect foresight of agents. Finally, the results are assessed and correctness of unconventional monetary policy is evaluated.

Keywords: unconventional monetary policy, foreign exchange interventions, occasionally binding constraint

JEL classification: E58
AMS classification: 91B64

1 Introduction

Until recently, the main monetary policy instrument in countries under inflation targeting regime were open market operations. Central banks performed expansionary or contractionary policy, to achieve steady price level rise, i.e. inflation at target value. Following the financial crisis in 2008, central banks were forced to drop the interest rate, in order to preserve steady growth of price level. Many central banks however, declined their interest rates to zero, i.e. they hit the zero lower bound. Few countries, as is the case of Czech Republic, started to intervene on exchange rate as a form of unconventional monetary policy. The aim was to depreciate the domestic currency, in order to increase inflation. This paper investigates on consequences of using interventions on foreign exchange market to meet inflation goal. Section 2 describes used model, section 3 introduces the way to model constraints on variables. Model estimation is described in section 4. Finally, section 5 shows comparison of economy with a constraint only on interest rate and the one with an additional one on exchange rate. Results are summarized in conclusion.

2 The model

We use New-Keynesian dynamic stochastic general equlibirum model of a small open economy under inflation targetting regime, described in [1]. The Euler equation is expressed as

$$c_t = \frac{1}{1+\upsilon} E_t c_{t+1} + \frac{\upsilon}{1+\upsilon} c_{t-1} - \frac{1-\upsilon}{\sigma(1+\upsilon)}[i_t - E_t \pi_{t+1} + \mu_t^d] \tag{1}$$

Phillips curve for domestic inflation is defined as:

$$\pi_{h,t} = \frac{\delta}{1+\delta\beta}\pi_{h,t-1} + \frac{\beta}{1+\delta\beta} E_t \pi_{h,t+1} + \frac{(1-\theta_h)(1-\theta_h\beta)}{\theta_h(1+\delta\beta)} mc_t \tag{2}$$

Monetary policy is modeled in usual way, using Taylor rule:

[1]Faculty of Economics and Administration, Masaryk University, Lipova 41a, Brno, Czech Republic, 405337@mail.muni.cz

[2]Faculty of Economics and Administration, Masaryk University, Lipova 41a, Brno, Czech Republic, osvald@econ.muni.cz

$$i_t = \rho_i i_{t-1} + (1-\rho_i)(\phi_\pi(\pi_t - \pi^T) + \phi_p P_t + \phi_y(y_t)) + \epsilon_t^i \tag{3}$$

Foreign economy is modeled as the rest of the world in a simpler manner. We implement price and nominal wage rigidities. To model exchange rate, we use modified uncovered interest parity condition described by [2]:

$$\tilde{E}_t e_{t+1} - e_t = i_t - i_t^* + \tau\sigma_e^2(\tilde{w}_t^* + \tilde{w}_t^{*,cb}), \tag{4}$$

where τ is the coefficient of absolute risk aversion of dealers on exchange market.. Capital inflows process $\tilde{w}_t^*$ is modelled as an AR(1) process. The above mentioned allows central bank to intervene on FX market through $\tilde{w}_t^{*,cb} = \chi_{e^T}(e_t - e^T) + \epsilon_t^{cb}$.

3 Occasionally binding constraints

Occasionally binding constraints (OBC) are modelled according to [3], by adding a shadow shock to the bounded variable. Suppose we have a vector of model's variables x_t and we want to know the response of these variables to shock ϵ_t. We also suppose that the model's equations are all linear, with the exception of variable $x_{1,t}$, which can be written as:

$$x_{1,t} = max\{0, \mu_1 + \phi_{-1}x_{t-1} + \phi_0 x_t + \phi_1 \mathbf{E}x_{t+1} - (\phi_{-1} + \phi_0 + \phi_1)\mu\}, \tag{5}$$

where $\mu = [\mu_1, .., \mu_n]$ is vector stacking the variables' steady states value and $\mu_1 > 0$. If the variable is driven to the bound, the shadow price shock will make the variable higher for some duration. Let IRF vectors be of length T and the number of periods, after which the constraint is no longer expected to bind, be denoted T^*. We assume $T \geq T^*$. The algorithm then replaces the aforementioned equation with:

$$x_{1,t} = \mu_1 + \phi_{-1}x_{t-1} + \phi_0 x_t + \phi_1 \mathbf{E}x_{t+1} - (\phi_{-1} + \phi_0 + \phi_1)\mu + \sum_{s=0}^{T^*-1} \epsilon_{s,t-s}^{SP}, \tag{6}$$

where $\epsilon_{s,t}^{SP}$ is the aforementioned shadow price shock. The algorithm then finds solution that binds variable at given value. It is also possible to consider more binded variables. The algorithm is same with generalization, based on adding more constraints. Also, stochastic simulation follows the same algorithm, i.e. adding shadow price shocks to variables at the bound. For details on algorithm, see [3].

4 Model estimation

4.1 Data

To estimate the model, we use 8 quarterly measured time series from CNB. The domestic economy is modeled as the case of Czech Republic. The Eurozone is considered as the foreign economy, being sufficiently large. The considered variables are domestic and foreign output measured as respective GDPs. Domestic and foreign inflation, measured as incremenets in CPI. Domestic and foreign interest rates, measured as 3month PRIBOR and EURIBOR respectively. Lastly, we use nominal exchange rate measured in CZK/EUR and Czech target inflation. Variables used are from first quarter of 2000 until last quarter of 2016, i.e. 68 observations. We transform the variables as followed: Outputs, inflation and exchange rate are measured as demeaned increment of differences of logarithms of respective variables. For domestic output, foreign output and exchange rate, we work with state variables, where for example $y_t = y_{t-1} + y_{obs}$, where y_{obs} is the demeaned increment. Interest rates are demeaned and divided by four, as to represent to represent the interest gained from one period to another. Target inflation is transformed in the same way to represent target inflation of respective quarters.

4.2 Parameters

The parameters are estimated using Bayesian techniques, namely Random Walk Chain Metropolis-Hastings algorithm, generating 1 000 000 draws in two chains, from which first 40% were burned. Calculatons were done using Matlab 2016b and Dynare 4.4.3. Seven parameters are calibrated. Discount factor β is set to 0.99. Following majority of literature, habit in consumption is set to 0.7 and share of import to 0.5, as is often used for case of Czech Republic. Parameters in the modified UIP are set following [4]: dealers' risk aversion is st to 500 and depreciation's variance is set to 0.028. The target exchange rate of central bank is set to exchange rate's steady state. Foreign inflation target is set to 0.38. The other parameters are estimated, as described above. Risk aversion parameters are estimated to be 1.35 and 1.30 for doomestic and foreign economy respectively, what shows rather risk averse consumers. Calvo parameter for wages was estimated at 0.75, meaning wages are updated yearly, and wage

indexation to past inflation parameter is 0.49. Prices Calvo is higher, estimated at 0.84 for both domestic goods and imports. Monetary rules are in line with expectations, inflation parameters being 1.49 and 1.54 and output parameters 0.5 and 0.31 for domestic and foreign economy respecitvely. Monetary policy smoothing parameters are estimated high, 0.97 and 0.9. Standard deviation of shocks are displayed in table 1.

Shock	Description	Prior mean	Posterior mean	90% HPD interval	Distribution	Prior std.
ϵ_a	technology	0.500	0.2401	[0.1174 , 0.3621]	invg	2.000
ϵ_d	demand	0.500	0.0990	[0.0757 , 0.1225]	invg	2.000
ϵ_w	wage	0.500	0.3624	[0.1223 , 0.6316]	invg	2.000
ϵ_p	cost-push	0.500	0.2298	[0.1143 , 0.3457]	invg	2.000
ϵ_r	monetary	0.010	0.0013	[0.0012 , 0.0014]	invg	0.100
ϵ_{cb}	cap. inflows	0.100	0.0243	[0.0178 , 0.0307]	invg	0.400
ϵ_d^*	f. demand	0.500	0.0697	[0.0603 , 0.0780]	invg	2.000
ϵ_a^*	f. technology	0.500	0.1504	[0.0949 , 0.2052]	invg	2.000
ϵ_r^*	f. monetary	0.010	0.0014	[0.0012 , 0.0015]	invg	0.100
ϵ_w^*	f. wage	0.500	0.2226	[0.1114 , 0.3306]	invg	2.000
ϵ_f^*	intervention	0.050	0.0128	[0.0093 , 0.0162]	invg	0.100
ϵ_{targ}	target	0.001	0.0003	[0.0003 , 0.0004]	invg	0.001

Table 1: Estimation of standard deviation of shocks

5 Stochastic simulation

We are next to assess the implications of unconventional monetary policy. First, we show behavior of economy at ZLB. Next, we show how constraint on nominal exchange rate changes such behavior. All computations were done using toolbox [5]. To assess the consequences this constraint has on an economy, we replace the model's monetary rule, with the following, non-linear rule:

$$i_t = max\{-0.005, \rho_i i_{t-1} + (1-\rho_i)(\phi_\pi \pi_t + \phi_p P_t + \phi_y y_t) + \epsilon_t^i\}. \tag{7}$$

The value -0.005 corresponds to ZLB, if mean of interest rate is modeled as steady state.

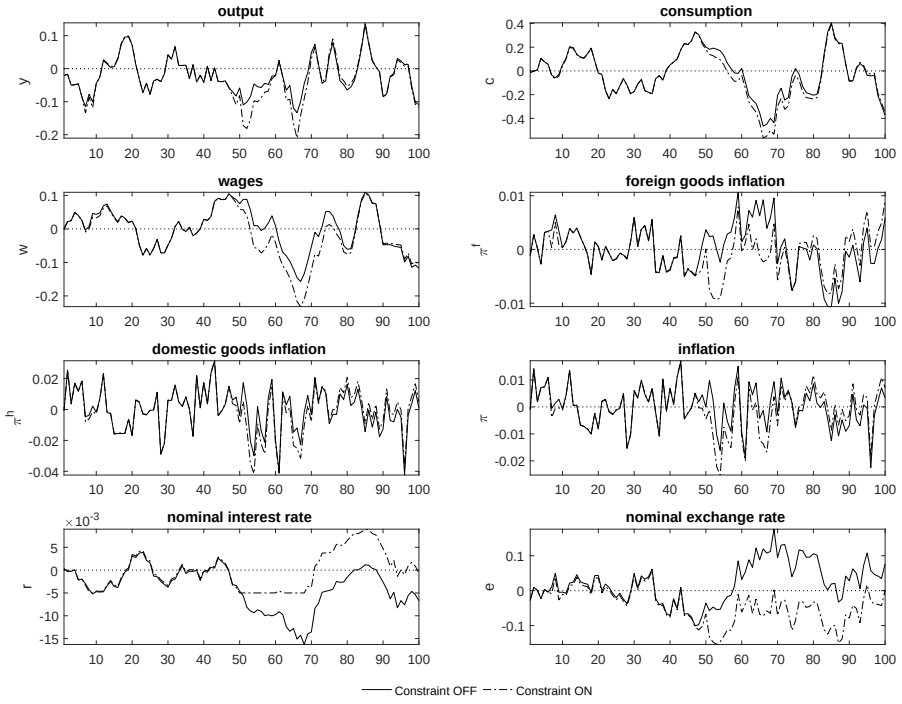


Figure 1: Simulation of economy at ZLB

Figure 1 shows simulations of chosen variables. The variables are depicted in states. The full lines represent the original simulation of the economy, i.e. the economy without ZLB constraint. Dashed lines therefore represent the economy, which is constrained by ZLB. It is seen, that nominal interest rate is constrained at two periods, first time shortly around period 10 and second time for longer time from period 50. If the economy spends only short period of time at the bound, it has only very small impact on the other variables. However, if we study the behavior of variables, when economy hits ZLB for longer time, we see that consumption decreases relative to the original trajectory. This effect was expected, as individuals face higher interest rate. Nominal exchange rate strongly appreciates. This is the conclusion of modified uncovered interest parity condition, where if interest rates is higher relative to original values, exchange rate must appreciate, as agents will look to invest in domestic economy. Economy's output is decreased as well, partly because of decreased consumption and partly because of appreciation of domestic currency. Consequently, wages decrease as well.

Next, we see, that domestic inflation decreases at ZLB as compared to original trajectory. This can be attributed to lower wages and lower terms of trade, meaning that the marginal costs of firms decrease. More significant difference between trajectories is seen in imported inflation. As exchange rate appreciates because of ZLB, importing firms lower their prices. The overall inflation in domestic economy is combined in equal parts from these two variables. It will therefore be lower, when economy is constrained, than it would be, if it was not. Interest rate at ZLB therefore ceases to be effective monetary policy instrument, and deflationary pressures may arise.

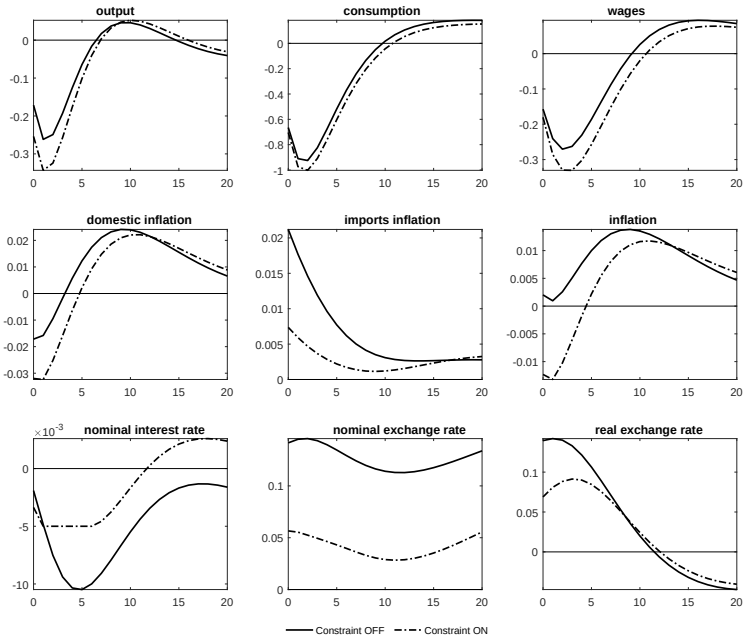


Figure 2: Demand shock

Next, we use impulse response functions to exogenous shocks to assess thebehavior of economy. These shocks are scaled by factor of 10, to make the differences clear. Figure 2 depicts demand shock, which is modeled as exogenous decrease in domestic consumption. Output, wages as well as consumption will decrease more in constrained economy. The feature of simulation is shown in IRF for interest rate, where in economy with constraint, it can not decrease as much, as a reaction to decreased output. Imports inflation will be lower, because exchange rate did not depreciate as much. In effect, inflation in constrained economy will be significantly lower, thanin unconstrained one. Nominal exchange rate depreciates less, because nominal interest rate is higher in constrained economy, causing appreciation pressures on nominal exchange rate.

In figure 3, we can examine, what happens, if an economy is hit by negative monetary shock, i.e. exogenous decrease in interest rate. Nominal interest rate only decreases to -0.005, which is the lower bound, as specified above. As interest rate remains higher, nominal exchange rate will not depreciate as much. On the other hand, because interest rate is higher, consumption will be lower than in the original case, because agents will be motivated to save, rather than spend. Both of these efects, smaller depreciation and smaller consumption will lead to smaller output. Imports inflation will not increase as in the case without the bound, because the exchange rate did not depreciate as much, and they have no reason to increase prices as much. To sum up, if the economy isn't able to decrease interest rates, output and inflation may remain low, and the economy can run into stagflation.

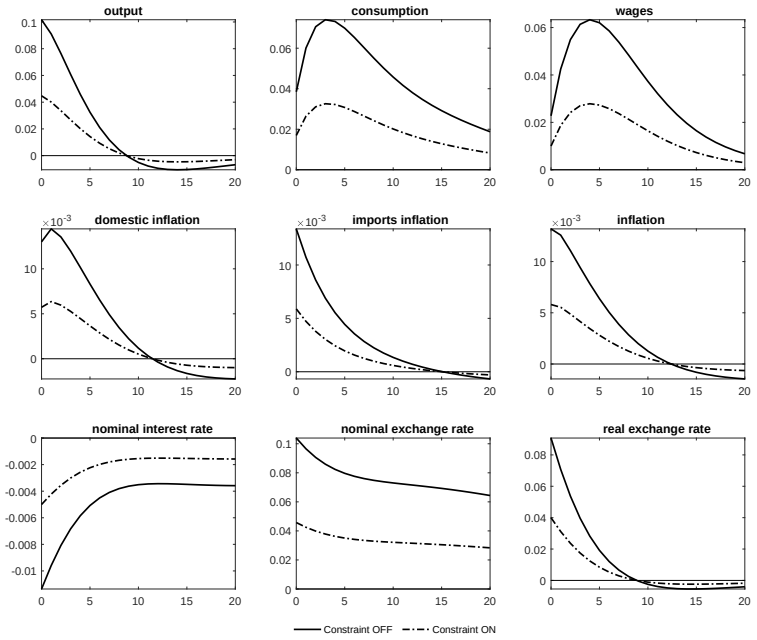


Figure 3: Negative monetary shock

5.1 Unconventional monetary policy

We can employ second constraint on the economy, namely a commitment on exchange rate employed by central bank. The constraint on interest rate is left as in the previous section. Next, we need to create a constraint, prohibiting exchange rate to appreciate. We use the modified uncovered interest parity condition, and use maximum function, providing, that exchange rate is either going to be the original value, or the value, under which it can appreciate. To determine the value, we look at the simulated economy above, and we choose approximate value -0.1, which we plug in the function to obtain:

$$e_t = max\{-0.1, E_t e_{t+1} + i_t^* - i_t - \tau\sigma_e^2(\tilde{w}_t^* + \tilde{w}_t^{*,cb})\} \tag{8}$$

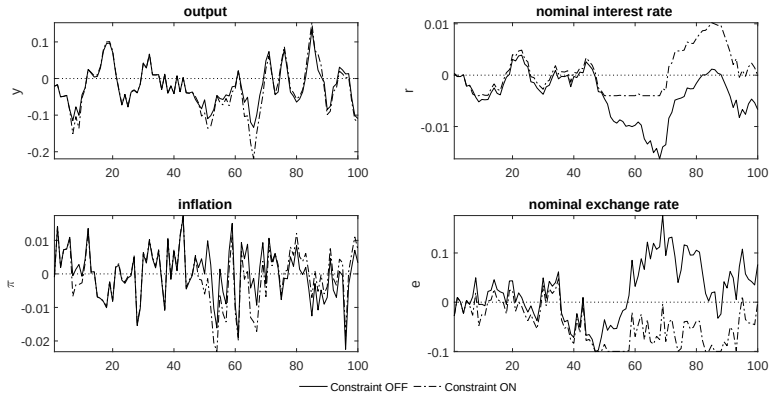


Figure 4: Simulation of economy at ZLB and Commitment

Figure 4 displays the same simulation as in the case above with additional constraint on exchange rate. We can observe how trajectory changes, when this constraint is employed. We see, that opposed to the case with single constraint, output does not decrease as much around period 50, i.e. when both bounds are in effect. Nominal interest rate still bounds at the same horizon. The effect of inflation is hard to assess and seems insignificant. Finally, nominal exchange rate is binded at -0.1 as described above. After it leaves the bound, it rises to slightly higher values, but this result is also not significant.

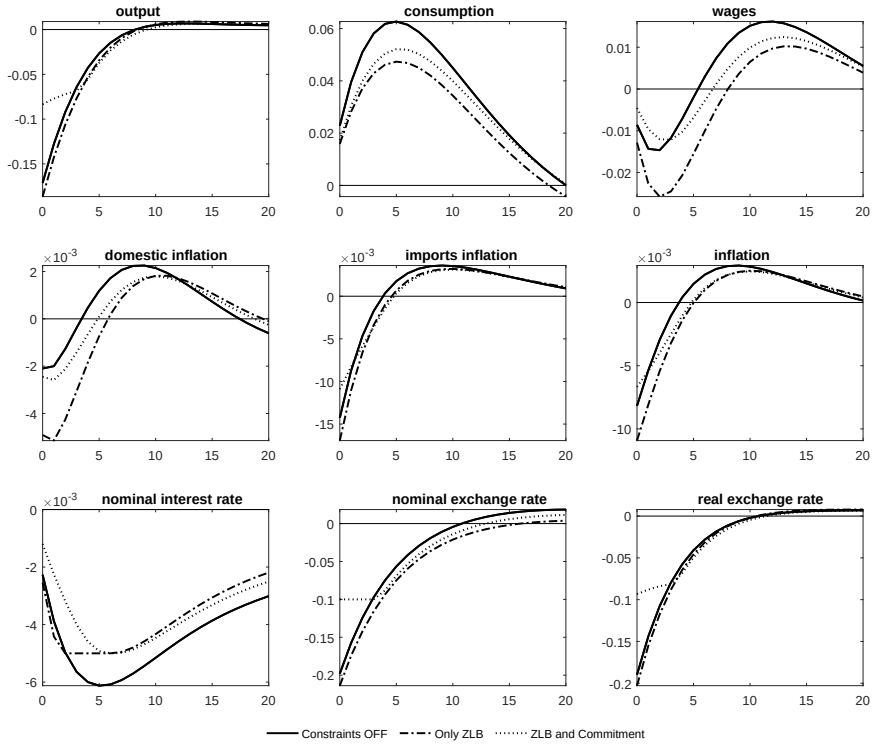


Figure 5: IRF to capital inflows shock

We can note, that behavior of the economy as response to aforementioned shocks does not change. Both demand shock and negative monetary shock will tend to depreciate the currency. Therefore the constraint isn't binding and the economy will reponse as described above. However, we can study the behavior of the economy under capital inflows shock. Such shock will have a negative effect on exchange rate, i.e. it will appreciate the domestic currency. However, with constraint in place, this will not be fully possible. Figure 5 examines the behavior of economy without any constraint, with ZLB constraint and both ZLB and commitment to certain exchang rate constraint.

We observe, that economy being bound by nominal exchange rate, will keep its exchange rate at the constrained. This will cause output to be higher as opposed to alternative without exchange rate constraint. Higher output will cause higher nominal interest rate. Imports inflation will be higher, at least in the short run, as well as domestic inflation, driving real interest rate down, and increasing consumption.

6 Conclusion

In this paper, we use non-linear New-Keynesian model to show, that if the economy is at zero lower bound, interventions preventing appreciation of currency have positive effect on economy. Such interventions by central bank tend to increase inflation, output and wages in an economy.

References

[1] Alpanda, S., Kotzé, K. and Woglom G.: The role of the exchange rate in a New Keynesian DSGE model for the South African economy. In: *South African Journal of Economics 78.2 (2010): 170-191.*

[2] Montoro, C. , and Ortiz, M.: Foreign exchange intervention and monetary policy design: a market microstructure analysis. *Third draft (2013). 119–124.*

[3] Holden, T., and Paetz, M.: Efficient simulation of DSGE models with inequality constraints. *School of Economics Discussion Papers 1612 (2012).*

[4] Malovaná, S.: Foreign exchange interventions at the zero lower bound in the Czech economy: A DSGE approach. In: *No. 13/2015. IES Working Paper, 2015.*

[5] Holden, T. (2016). DynareOBC v1.2.0957 [Data set]. Zenodo. *http://doi.org/10.5281/zenodo.50132*

Spatial panel data models - stability analysis with application to regional unemployment

Tomáš Formánek[1], Roman Hušek[2]

Abstract. Regional macroeconomic policy actions should be prepared, implemented and evaluated while accounting for the spatial nature of their effects. During the quantitative analysis process, regional distances, interactions and spill-overs need to be properly addressed. Over the last few years, panel data-based methods of spatial analysis have gained considerable attention as panels provide many advantages over cross-sectional spatial data in terms of tackling the temporal aspects of the macroeconomic dynamics as well as by allowing to account for unit's individual effects.
In this article, we focus on selected key topics related to the estimation and stability of spatial panel models. We estimate and evaluate a panel data-based model describing regional unemployment dynamics at the NUTS2 level for the following six countries: Czechia, Slovakia, Poland, Hungary, Germany and Austria. Different methodological approaches are used to study key aspects of the spatio-temporal stability of the estimated model. Based on our results, we may draw reasonably general conclusions with respect to empirical regional unemployment policies.

Keywords: spatial panels, model stability, regional unemployment.

JEL classification: C23, C31, C52, E66
AMS classification: 91B72

1 Introduction

Spatial econometric models address the presence of effects such as economic spill-overs between neighboring regions. Spatial models play an important role in regional studies and in many non-economic fields of research. In spatial econometrics, data need to be geo-coded by means of the latitude/longitude geographic coordinates system, as distances (and/or common borders) are used to estimate spatial dependencies. The variety of available approaches towards modeling and estimation of spatial dependencies imply that researchers usually have to consider several spatial structure settings to evaluate coefficient stability and model robustness. The analysis of spatial panel data is an important field of contemporary econometrics. Spatial econometrics contributes or may significantly contribute to many fields of current econometric research, either explicitly or implicitly, see e.g. [2], [3], [4], [10] or [11] for empirical and theoretical contributions to the field as well as for additional references.

In this paper, we use spatial panel data methods to evaluate the conditional stability of estimated spatial panel models. The theoretical part of our analysis is accompanied by a panel data-based spatial model describing regional unemployment dynamics at the NUTS2 level for the following six countries: Czechia, Slovakia, Poland, Hungary, Germany and Austria. For basic illustration of the spatio-temporal dynamics and dependencies in regional unemployment, Figure 1 is provided.

The remainder of this paper is structured as follows: Section two covers selected key topics of the spatial panel approach and provides references to fundamental literature. Section three provides an illustrative application to the topics outlined. Section four and the list of references conclude our paper.

2 Spatial econometrics & panel data models

The actual framework for contemporary applied spatial econometrics was provided by Cliff and Ord (e.g. in [1]), by introducing a flexible spatial matrix $\boldsymbol{S}_N$ that is used to define neighbors (spatially close regions) using a dummy variable technique, where each element of the square spatial matrix equals 1 if the two spatial units are neighbors and 0 otherwise. Usually, two units are considered neighbors if they are close in some adequate dimension (geographic, technological, etc.).

A corresponding spatial weights matrix $\boldsymbol{W}_N$ is often constructed by simple row-standardizing the spatial matrix, so that the row weights sum up to 1. A simple 4-unit (4×4) example is provided in (1). From the first row (and column) of the symmetric spatial matrix $\boldsymbol{S}_N$, we may observe that the first unit (say, region or city) is a neighbor of units 2, 3 and 4. Diagonal elements are set to zero by definition (units are not neighbors to themselves). Clearly,

[1]VŠE Praha, nám. W. Churchilla 4 Praha 3, Prague 130 67, Czech Republic, formanek@vse.cz
[2]VŠE Praha, nám. W. Churchilla 4 Praha 3, Prague 130 67, Czech Republic, husek@vse.cz

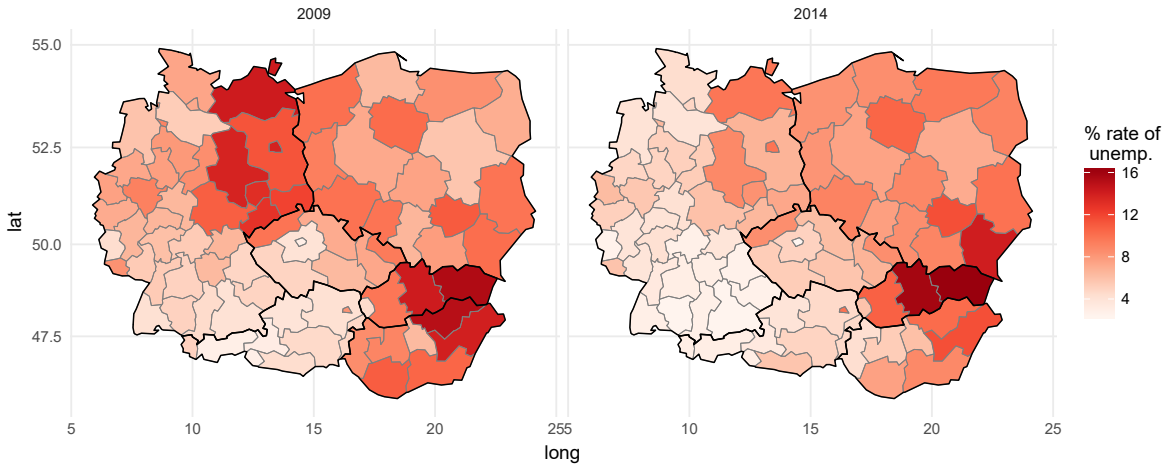


Figure 1 Choropleths with 2009 & 2014 unemployment rates at the NUTS2 level

both $\boldsymbol{S}_N$ and $\boldsymbol{W}_N$ matrices are determined by the ordering of the data, which can be arbitrary. Elhorst [4] provides two formal stability conditions for spatial models that may be restated as follows: (a) The row and column sums of any $\boldsymbol{S}_N$ matrix should be uniformly bounded in absolute value as $N \to \infty$. (b) The row and column sums of $\boldsymbol{S}_N$ should not diverge to infinity at a rate equal to or faster than the rate of sample size growth.

$$\boldsymbol{S}_4 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{W}_4 = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}. \tag{1}$$

As we take advantage of the notation of Millo and Piras [7], a general form of a static panel model that includes both the spatial effects (spatial lag) of the dependent variable and the spatially autocorrelated error terms may be outlined as follows:

$$\boldsymbol{y} = \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W}_N \right) \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \tag{2}$$

where $\boldsymbol{y}$ is a $(N \times T) \times 1$ column vector of dependent variable observations ($i = 1, 2, \ldots, N$ denotes cross-sectional units and $t = 1, 2, \ldots, T$ relates to the time dimension). $\boldsymbol{X}$ is a $(N \times T) \times k$ matrix of k exogenous regressors, has full column rank and its elements are uniformly bounded in their absolute values. $\boldsymbol{I}_T$ is an identity matrix and $\boldsymbol{W}_N$ is a time-invariant spatial weights matrix describing the geographical arrangement of the sample units ($\boldsymbol{W}_N$ conforms to the conditions described above). Elements of vector $\boldsymbol{\beta}$ as well as λ are parameters of the model. Kronecker product (direct matrix product) is used and denoted $\otimes$. Given the space limitations for this contribution, we only focus on a random effects (RE) model/estimation approach. As we do not elaborate on the pooling nor fixed effects (FE) approaches, we refer the readers to [4] and [7] for detailed discussion of the topic. With RE models, we implicitly assume that the unobserved individual effects are not correlated with other regressors. Testing of random effects in an estimated model (RE vs. FE tests) is an essential part of RE model verification. Although such specification tests lie beyond the scope of this paper, they are readily available from [7] as well as implemented into R software: {spdep} and {splm} packages. Hence, using the notation as in [7], the error term $\boldsymbol{u}$ from (2) and its variance may be described as

$$\boldsymbol{u} = \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N \right) \boldsymbol{\mu} + \left(\boldsymbol{I}_T \otimes \boldsymbol{B}_N^{-1} \right) \boldsymbol{\upsilon}, \tag{3}$$

$$\boldsymbol{\Omega}_u = var\left(\boldsymbol{u} \right) = \sigma_\mu^2 \left(\boldsymbol{\iota}_T \boldsymbol{\iota}_T' \otimes \boldsymbol{I}_N \right) + \sigma_\upsilon^2 \left[\boldsymbol{I}_T \otimes \left(\boldsymbol{B}_N' \boldsymbol{B}_N \right)^{-1} \right], \tag{4}$$

where $\boldsymbol{\iota}_T$ is a unit vector $(T \times 1)$ and $\boldsymbol{I}_N$ is an $(N \times N)$ identity matrix. Vector $\boldsymbol{\mu}$ holds the time-invariant and spatially uncorrelated individual effects with $\mu_i \sim IID\left(0, \sigma_\mu^2\right)$. $\boldsymbol{B}_N = \left(\boldsymbol{I}_N - \rho \boldsymbol{W}_N \right)$ is assumed non-singular and features a spatial error autoregression parmater ρ where $|\rho| < 1$. $\boldsymbol{\upsilon}' = \left(\boldsymbol{\upsilon}_1', \ldots, \boldsymbol{\upsilon}_T' \right)$ is a vector of innovations that vary both over cross-sectional units and across time with $\upsilon_{it} \sim IID\left(0, \sigma_\upsilon^2\right)$.

Maximum likelihood (ML) or generalized moments (GM) procedures implemented into R software allow us to estimates the $\boldsymbol{\beta}$, λ and ρ parameters, along with σ_v^2 and σ_μ^2. For a detailed description and R implementation of the estimation, see [7].

Beyond model estimation, it is crucial to to keep in mind some basic interpretation issues related to a model featuring a spatial lag in the dependent variable: The estimated parameters of the model (2) do not form proper basis for a description of model dynamics, especially if we want to focus on spillover effects. As we simulate a change in x_{rit} - the r-th explanatory variable for spatial unit i at time t - we expect the dependent variable in the i-th unit to change (direct effect) and also, for $\lambda \neq 0$, we expect some non-zero effects on the dependent variables in neighboring units (indirect effects). As we only deal with static models here, all effects take place at time t. Considerable efforts are made in current literature to introduce feasible dynamic spatial panel estimators - however, strong results are yet to be seen.

For proper direct and indirect effect estimation, we have to use partial derivatives of the regression function. Even with static spatial panel models, this type of dynamics is relatively complex to describe. However, we may use the notation of LeSage and Pace [6] to provide a relatively simple overview: We start with a reduced form of a cross-sectional spatial model:

$$(\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)\, \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \alpha \boldsymbol{\iota}_N + \boldsymbol{u}, \tag{5}$$

where $\boldsymbol{y}$ and $\boldsymbol{u}$ are $(N \times 1)$, $\boldsymbol{X}$ is $(N \times k)$ and α is the intercept. Equation (5) can be conveniently rewritten for subsequent interpretation as

$$\begin{aligned} \boldsymbol{y} &= \sum_{r=1}^{k} \boldsymbol{S}_r(\boldsymbol{W}_N)\, \boldsymbol{x}_r + \boldsymbol{V}_N(\boldsymbol{W}_N)\, \boldsymbol{\iota}_N \alpha + \boldsymbol{V}_N(\boldsymbol{W}_N)\, \boldsymbol{u}, \\ \boldsymbol{S}_r(\boldsymbol{W}_N) &= \boldsymbol{V}_N(\boldsymbol{W}_N)\, \boldsymbol{I}_N \beta_r = (\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)^{-1}\, \boldsymbol{I}_N \beta_r, \\ \boldsymbol{V}_N(\boldsymbol{W}_N) &= (\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)^{-1} = \boldsymbol{I}_N + \lambda \boldsymbol{W}_N + \lambda^2 \boldsymbol{W}_N^2 + \lambda^3 \boldsymbol{W}_N^3 + \ldots \end{aligned} \tag{6}$$

Essentially, the impacts (direct and indirect) are related to the expression

$$\boldsymbol{S}_r(\boldsymbol{W}_N) = (\boldsymbol{I}_N - \lambda \boldsymbol{W}_N)^{-1}\, \boldsymbol{I}_N \beta_r. \tag{7}$$

For a static spatial panel model environment, (7) may be generalized to

$$\boldsymbol{S}_r(\boldsymbol{\mathcal{W}}) = (\boldsymbol{I}_{NT} - \lambda \boldsymbol{\mathcal{W}})^{-1}\, \boldsymbol{I}_{NT} \beta_r. \tag{8}$$

Here, $\boldsymbol{\mathcal{W}} = (\boldsymbol{I}_T \otimes \boldsymbol{W}_N)$ is a block-diagonal matrix with diagonal blocks formed by the spatial matrices $\boldsymbol{W}_N$. From (6), we may see that the direct and indirect impacts for a cross-sectional spatial model are given by

$$\frac{\partial y_i}{\partial x_{ri}} = \boldsymbol{S}_r(\boldsymbol{W}_N)_{ii} \quad ; \quad \frac{\partial y_i}{\partial x_{rj}} = \boldsymbol{S}_r(\boldsymbol{W}_N)_{ij}\,, \tag{9}$$

where $\boldsymbol{S}_r(\boldsymbol{W}_N)_{ij}$ is a scalar term, element of the matrix $\boldsymbol{S}_r(\boldsymbol{W}_N)$. Now, we may conclude our derivation of impacts for model (2): to express the impacts from an estimated spatial panel model, we only need to substitute $\boldsymbol{S}_r(\boldsymbol{\mathcal{W}})$ for $\boldsymbol{S}_r(\boldsymbol{W}_N)$ in equation (9). For detailed discussion of impacts' variances and statistical significance tests, see [6].

Stability of an estimated RE spatial model

One major weakness of the spatial econometric approach as described above is the fact that $\boldsymbol{W}_N$ matrices cannot be estimated along with model parameters. Rather, $\boldsymbol{W}_N$ needs to be specified prior to model estimation. There is little theoretical background for choosing the *right* $\boldsymbol{W}_N$ specification. In [5], we discuss the consequences of possible misspecification of $\boldsymbol{W}_N$ within the scope of cross-sectional spatial models. Also, a simple yet effective algorithm is provided for choosing an optimal $\boldsymbol{W}_N$ from a group of alternative spatial matrices as in (2). In our contribution, different distance-based weights matrices are produced by varying maximum neighbor distance thresholds. This is more realistic for our unemployment-related illustrative application as compared to the k-nearest neighbors method. For details, please refer to a discussion provided in [5].

3 Empirical application

This section provides an estimation and evaluation of a relatively compact spatial panel model, rooted in the regional competitiveness theory (please refer to [5] and [9]). At the NUTS2 level, regional unemployment rates in Czechia, Slovakia, Poland, Germany, Austria and Hungary are modeled using relative GDP per capita and technological advantage indicators, along with other control variables (described below). Specifically, we have

82 NUTS2 regions: 8 regions in Czechia, 4 in Slovakia, 9 in Austria, 38 in Germany (8 are former East-German regions; 9 including the unified Berlin), 16 in Poland and 7 in Hungary. For those regions, we use a balanced (annual) 2009-2014 dataset with the following variables downloaded from Eurostat via R package {eurostat}: the "lfst_r_lfu3rt" dataset is the source of our geo-coded dependent variable (unemployment rate). GDP per capita (measured as percentage of the EU average: a competitiveness-based indicator) comes from "nama_10r_2gdp" and the "htec_emp_reg2" dataset is used to describe the percentage of employees working in the high-tech sector (NACE r.2 code HTC).

Hence, the equation (2) is estimated as follows: y_{it} is the total unemployment rate (percentage of labor force - ages 20-64 combined). For each spatial unit, the vector of regressors $\boldsymbol{x}'_{it}$ consists of the following six variables: 1: GDP_{it} is the relative GDP per capita (100 = EU average in a given year). 2: HTC_{it} is the high-tech sector employment proportion. 3: $HUCl_{it}$ is a dummy variable based on Ord's G^* (see [8]) that discerns local clusters (hotspots) of high values of the observed variable. As we searched for high unemployment clusters, we found a single cluster, containing the following NUTS2 regions: HU10, HU31, PL21, PL22, PL32, PL33, SK03 and SK04 (generally speaking, this cluster consist of units close to or bordering with Ukraine). 4: $Intrc_{it} = HTC_{it} \times HUCl_{it}$ is an interaction term that controls for differences in effects of high-tech sector prominence in regular vs. high-unemployment regions. At the regional level, this particular variable has a strong policy-making and policy evaluation potential in terms of tackling high unemployment through high-tech oriented subsidiary measures, etc. 5: $\overline{GDP}_i$ and 6: $\overline{HTC}_i$ are the individual (time-invariant) averages, necessary for estimation of the $\boldsymbol{\beta}$ coefficients using a correlated random effects-like (CRE) approach as described in Wooldridge [12].

After applying basic spatial correlation tests to the observed data and verification of additional model assumptions (see [5]), we estimate the model (2). Table 1 provides the estimated coefficients along with their standard errors and significance levels - based on a $\boldsymbol{W}_{82}$ matrix defined for maximum neighbor distances of 240 km. This spatial weights specification leads to the highlest log-likelihood of the estimated regression function. The estimated equation (as presented in Table 1) is statistically significant at the 5% level and the estimated residuals are not spatially correlated.

	Estimate	Std. Error	t-value	Pr(>|t|)
(*const*)	2.401	0.532	4.516	0.000
GDP	-0.043	0.013	-3.284	0.001
HTC	0.226	0.085	2.645	0.008
HUCl	2.639	0.949	2.781	0.005
Intrc	-0.377	0.215	-1.754	0.080
$\overline{GDP}$	0.032	0.014	2.364	0.018
$\overline{HTC}$	-0.301	0.151	-1.987	0.047
λ	0.820	0.033	25.055	0.000

Table 1 Coeffs for neighbor distance threshold set at 240 km

As we have discussed in section 2, the estimated $\boldsymbol{\beta}$ coefficients should not be used for model dynamics description, i.e. they do not represent the ceteris paribus expected effects of a change in a given regressor (say, a r-th regressor x_{rit}) on the dependent variable for the i-th region and in the neighboring regions (we do not expect y_{it} in non-neighboring regions to be affected by changes in x_{rit}). Hence, Table 2 is provided, featuring both the direct and indirect effects along with their significance levels. (see [4] for a detailed discussion).

Effects	*GDP*	*HTC*	*HUCl*	*Intrc*	$\overline{GDP}$	$\overline{HTC}$
Direct	-0.049	0.259	3.115	-0.439	0.037	-0.346
z-score	-3.163	2.635	2.894	-1.803	2.289	-1.920
p-value	0.002	0.008	0.004	0.071	0.022	0.055
Indirect	-0.194	1.025	12.325	-1.742	0.148	-1.369
z-score	-2.529	2.171	2.288	-1.606	1.995	-1.685
p-value	0.011	0.030	0.022	0.108	0.046	0.092

Table 2 Impacts for neighbor distance threshold set at 240 km

The information provided in Table 2 can be summarized as follows: given a one percent increase in region's relative GDP per capita (a relative indicator with competitiveness-based interpretation), we can expect a mild, yet

statistically significant decrease in the expected rate of unemployment. Here, the expected spillover effect is about $4\times$ higher than the direct effect, which might be seen as a strong argument in favor of regional macroeconomic co-operation. The impacts of $HUCl_{it}$ have a rather technical interpretation: this variable is used to control for unemployment dynamics specificities in the empirically observed high unemployment cluster. The same technical interpretation applies to $\overline{GDP}_i$ and $\overline{HTC}_i$ variables of the CRE-like specification used for model (2). The direct and indirect effects of the variable HTC_{it} are rather interesting, as they differ in sign when calculated for standard regions vs. high-unemployment cluster regions: from our results, it seems that any potential subsidiary measures aimed at the high-tech sector would only have counter-unemployment effect for regions in the high-unemployment cluster. The λ coefficient as in Table 1 indicates that the estimated spatial dependence is rather strong and highly statistically significant.

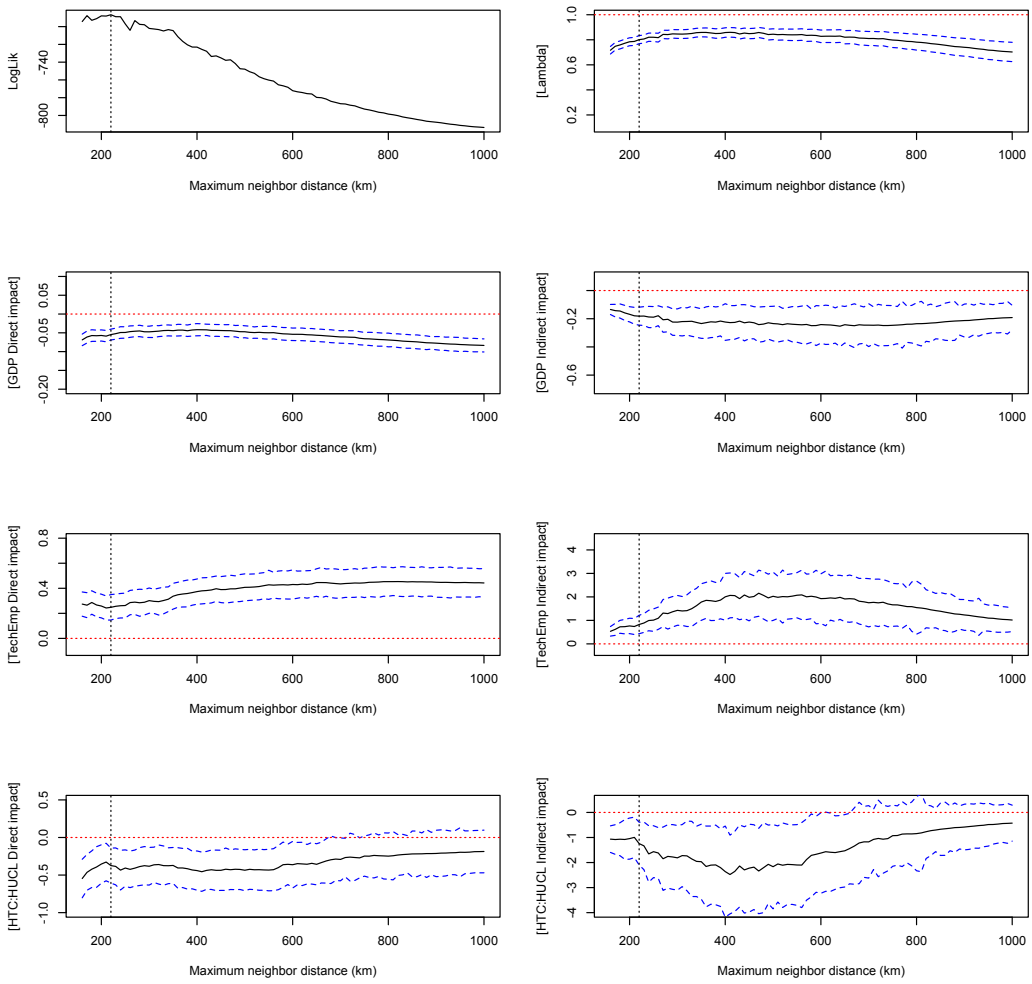


Figure 2 Stability analysis of the spatial panel model

Also, given the prevailing macroeconomic differences between the "old" and "new" EU countries and regions, we have considered an additional dummy variable for our model (2), distinguishing the old EU regions (Austrian and former West-German NUTS2 regions). Interestingly, once the $HUCl_{it}$ variable and its interaction term are controlled for, such additional dummy variable (its estimated coefficients and impacts) turns out statistically insignificant, even at the $\alpha = 0.1$ level.

To account for the variety of possible $\boldsymbol{W}_{82}$ specifications and the resulting ambiguity potential, specification robustness of equation (2) was tested against changes in neighborhood definitions. Multiple estimations of model (2) were performed, while varying the spatial weights matrix $\boldsymbol{W}_{82}$. The process of model-robustness evaluation is summarized in Figure 2: We start with a very sparse $\boldsymbol{W}_{82}$, constructed using maximum neighbor distance threshold set to 160 km (lower distances generate "islands" that are incompatible with the maximum likelihood estimation of spatial models). Next, neighbor thresholds are iteratively increased (using 10 km iterations, 85 models are evaluated) and new $\boldsymbol{W}_{82}$ matrices are generated - up to a maximum neighbor distance of 1.000 km (beyond 1.000 km, our spatial model provides no significant improvement over OLS, as discussed in [5]). Figure 2 shows the

relative stability of the estimated direct impacts and spillovers (along with their asymptotic $\pm$ 1 s.e.) over a relatively large span of maximum neighborhood distances. The top-left Figure 2 item shows the log likelihoods of the individual estimated models. This information was used to select our "best" specification for Tables 1 and 2 and their interpretation.

4 Conclusions

Spatial panel models provide an analysis framework where spatio-temporal aspects and patterns can be controlled for when quantifying and interpreting regional macroeconomic dynamics. Spatial panel models have the ability to discern between geographical determination and the influence of relevant macro-economic variables, many of which may be subject to or directly controlled by economic policy actions undertaken by the central authorities at different levels.

Our results support the importance of regional cooperation in macroeconomic policies addressing unemployment. Also, strong spatial autocorrelation was identified in the observed unemployment data. This study also provides model stability evaluation that supports model robustness and its high potential towards the analyses of dynamics in related fields of macroeconomic research. Spatial panel models provide a modern, useful, interpretable and functional approach towards regional (macroeconomic) data analysis.

Acknowledgements

Geo-data source: GISCO - Eurostat (European Commission), Administrative boundaries: © EuroGeographics. Supported by the grant No. IGA F4/73/2016, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Cliff, A. D., and Ord, J. K.: *Spatial Processess: Models and Applications*. Pion, London, 1981.

[2] Černý, M., Hladík, M.: Two optimization problems in linear regression with interval data. *Optimization* **66**(3), (2017), 331–349.

[3] Čížek, O.: Makroekonometrický model eurozóny, *Politická ekonomie* **63**(3), (2015), 279—299.

[4] Elhorst, J. P.: *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer-Briefs in Regional Science, Berlin, Springer, 2014.

[5] Formánek, T., and Hušek, R.: On the stability of spatial econometric models: Application to the Czech Republic and its neighbors. In: *Mathematical Methods in Economics* (A. Kocourek, and M. Vavroušek, eds.), TU Liberec, Liberec, 2016, 213–218.

[6] LeSage, J. P., and Pace, R. K.: *Introduction to spatial econometrics*. CRC Press, Taylor & Francis Group, Boca Raton, 2009.

[7] Millo, G., and Piras, G.: splm: Spatial panel data models in R, *Journal of Statistical Software* **47**(1), (2012), 1—38.

[8] Ord, J. K., and Getis, A.: Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis* **27** (1995), 286—306.

[9] Philipp, M.: Agglomeration and FDI in East German Knowledge-Intensive Business Services, *Economia politica* **3** (2012), 343–360.

[10] Tahal, R, Stříteský, V.: Building the Effective Loyalty Program. In: *CONFENIS-2013* (J. Basl, et al., eds.), Trauner Verlag, Linz, 2013, 251—258.

[11] Tepperová, J., Zouhar, J., and Wilksch, F.: Intra-EU Migration: Legal and Economic View on Jobseekers' Welfare Rights, *Journal of International Migration and Integration* **17** (2016), 1—20.

[12] Wooldridge, J. M..: *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, 2010.

ANP analysis and selection of the appropriate managerial methods

Veronika Frajtová, Helena Brožová[1]

Abstract. In recent years, the pressure from customers, purchasers and competitors forces companies to constant improvement in the strategic processes at the management level. The companies that are under increasing pressure try to look for modern management techniques that could improve and innovate their processes or products. Techniques that would lead to reductions in costs eliminate defects, increase efficiency and quality. All quality-focused methods and standards of management of quality and innovation help organizations to set-up whole management-system to prevent negative phenomena. However, what a method is the right one? Which do they should to pick from the ocean of methods and techniques? Difficult question. Every company certainly has its own know-how, best practices, proven methods, etc. Therefore, this paper proposes models based on the Analytic Network Process. ANP technique is structured multi-criteria evaluation approaches for organizing and analysing complex decisions as the determination of the managerial method that could be the best one for the specific economic sector.

Keywords: Analytic network process, process, the managerial methods.

JEL Classification: C44
AMS Classification: 91B50, 91B15

1 Introduction

Current modern period and demand from the business side caused that the companies with different specialization stuck in the cycle of the constantly improvements. This cycle pushes them into perpetually process innovation and evolves present methods and technologies to reach the competitive advantage on the market. Companies should be much more flexible and adaptable to customer's criteria.

The pressure from the competition forced them to strategical decision-making, which may influence the whole enterprise development and effectiveness. The final decision should be comprehensively and systemically effective based on the systems approach.

Nowadays, there exist a plenty of modern methodologies, methods and tools which will maximize the speed of flow of products or services with minimal defects.

The aim of this paper is to analyse suitability, importance or usability of modern managerial methods using the proper ANP model, which is quantified, based on a questionnaire survey among managers.

2 Materials and methods

2.1 Analytic Network Process

The Analytic Hierarchy Process (AHP) and its generalization to dependence and feedback, the Analytic Network Process (ANP), are methods of relative measurement of tangibles and intangibles. Being able to derive such measurements is essential for making good decisions (Saaty, 2009, 2012).

The analytic network process can model complex decision problems where a hierarchical model is not sufficient, because the problem cannot be shown in a hierarchy. This method can handle dependences and feedbacks among the components. It is a generalized version of analytic hierarchy process.

A network contains clusters and components in these clusters. The components in one cluster can be connected to components in another cluster (outer dependence) or the same cluster (inner dependence). A network is concerned with all the influences that can affect an outcome. The differences between a hierarchy and a network are shown in Figure 1.

[1] Department of Systems Engineering, Faculty of Economics and Management, Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic. frajtova@pef.czu.cz, brozova@pef.czu.cz

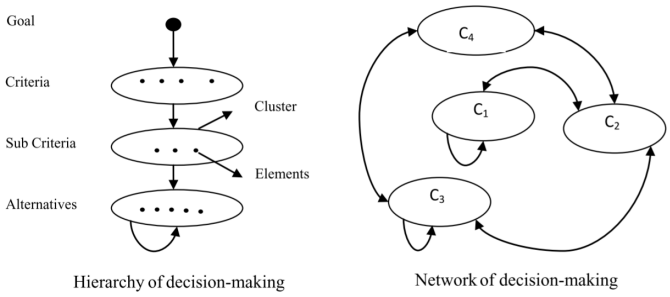


Figure 1 AHP and ANP structure differences

The evaluation of the dependencies in the network structure is organized in the super-matrix. Supermatrix for the network from the **Figure 1** is in the **Table 1**.

		C_1				C_2				C_3				C_4			
		a_{11}	a_{12}	...	a_{1n_1}	a_{21}	a_{22}	...	a_{1n_2}	a_{31}	a_{32}	...	a_{1n_3}	a_{41}	a_{42}	...	a_{1n_4}
C_1	a_{11}	w_{11}^{11}	w_{11}^{12}	...	$w_{11}^{1n_1}$	w_{11}^{21}	w_{11}^{22}	...	$w_{11}^{2n_2}$								
	a_{12}	w_{12}^{11}				w_{12}^{21}											
	...	...		...		...		...									
	a_{1n_1}	$w_{1n_1}^{11}$			$w_{1n_1}^{1n_1}$	$w_{1n_1}^{21}$			$w_{1n_1}^{2n_2}$								
C_2	a_{21}	w_{21}^{11}	w_{21}^{12}	...	$w_{21}^{1n_1}$					w_{21}^{31}	w_{21}^{32}	...	$w_{21}^{3n_3}$	w_{21}^{41}	w_{21}^{42}	...	$w_{21}^{4n_4}$
	a_{22}	w_{22}^{11}								w_{22}^{31}				w_{22}^{41}			
	...	...		...						...		...		...		...	
	a_{1n_2}	$w_{2n_2}^{11}$			$w_{2n_2}^{1n_1}$					$w_{2n_2}^{31}$			$w_{2n_2}^{3n_3}$	$w_{2n_2}^{41}$			$w_{2n_2}^{4n_4}$
C_3	a_{31}					w_{31}^{21}	w_{31}^{22}	...	$w_{31}^{2n_2}$	w_{31}^{31}	w_{31}^{32}	...	$w_{31}^{3n_3}$	w_{31}^{41}	w_{31}^{42}	...	$w_{31}^{4n_4}$
	a_{32}					w_{32}^{21}				w_{32}^{31}				w_{32}^{41}			
	...					...		...		...		...		...		...	
	a_{1n_3}					$w_{3n_3}^{21}$			$w_{3n_3}^{2n_2}$	$w_{3n_3}^{31}$			$w_{3n_3}^{3n_3}$	$w_{3n_3}^{41}$			$w_{3n_3}^{4n_4}$
C_4	a_{41}					w_{41}^{21}	w_{41}^{22}	...	$w_{41}^{2n_2}$	w_{41}^{31}	w_{41}^{32}	...	$w_{41}^{3n_3}$				
	a_{42}					w_{42}^{21}				w_{42}^{31}							
	...					...		...		...		...					
	a_{1n_4}					$w_{4n_4}^{21}$			$w_{4n_4}^{2n_2}$	$w_{4n_4}^{31}$			$w_{4n_4}^{3n_3}$				

Table 1 Super-matrix of the ANP structure from

The ANP can be implemented in the following consecutive steps:

1. Creating the problem network with clusters and its components.
2. Computing the vectors of weights of clusters according to the superior component using the Saaty's pairwise comparisons.
3. Computing the vectors of weights of components in each cluster according to the superior component using the Saaty's pairwise comparisons.
4. Creating the super-matrix, which represents the influence priority of a component on the left of the matrix on a component at the top of the matrix.
5. Calculation of weighted super-matrix, which columns sum is equal to 1.
6. Calculating of the limited super-matrix, which consists of the global preferences of decision components.

The final weights are the elements of the limited super-matrix and represent importance or preferences of the problem components from the main solved goal.

2.2 The Managerial Methods

This research attempted to analyse suitability, importance or usability of modern managerial methods. To provide this analysis the following broadly used methods were selected.

ITIL - The most widely used approach for IT service management ITIL provides a practical framework for planning, designing, delivering, supporting and improving IT services (OGC, 2010, Kiran, 2011). ITIL advocates that IT services must be aligned to the needs of the business and underpin the core business processes. It provides guidance to organizations on how to use IT as a tool to facilitate business change, transformation and growth. The goal of this strategic thinking is to define a plan that will bring a solution of the problem in a particular situation. It focuses on the value for customer and identifies the strategic benefits that can be used as a competitive advantage.

PRINCE2 is project management methodology developed by the government of the United Kingdom (UK) and used internationally, especially in information technology (IT) environment (Hinde, 2009). PRINCE2 makes use of the best-proven practices from a variety of industries and backgrounds. Documents with templates and clear decision points are characteristics of this methodology, which outlines eight processes.

Kanban is a visual tool for managing the flow of something or anything, like information or work items. It is flexible enough to be introduced in an existing process, whatever that process may be and without having the need to replace the process, but is very compatible with Agile methods. Kanban helps in the management of projects by visualizing workflow, limiting work in process, and creating and improving the flow. Kanban promotes continuous collaboration and encourages active, ongoing learning and improving by defining the best possible team workflow (Anderson, 2010).

Kaizen is a philosophy and practice that sees improvement in productivity as a gradual and methodical process (Imai, 2008). Kaizen is a Japanese term meaning "change for the better." The concept of Kaizen encompasses a wide range of ideas. It involves making the work environment more efficient and effective by creating a team atmosphere, improving everyday procedures, ensuring employee satisfaction, and making a job more fulfilling, less tiring and safer.

Lean management and **Six Sigma** - Lean management is often combined with Six Sigma tools (George and Lawrence, 2002, Töpfer, 2008). The Lean Six Sigma utilizes the DMAIC phases similar to that of Six Sigma. Lean Six Sigma projects comprise aspects of Lean focus on waste elimination and the Six Sigma focus on reducing defects based on the qualitative (CTQ) characteristics. The DMAIC method or model is more common. Every phase in this model has an important purpose and different procedures that are used to secure the correct results. The abbreviation DMAIC means Define, Measure, Analyse, Improve, and Control.

Total Quality Management (TQM) is an approach that seeks to improve quality and performance, which will meet or exceed customer expectations. This can be achieved by integrating all quality-related functions and processes throughout the company. TQM looks at the overall quality measures used by a company including managing quality design and development, quality control and maintenance, quality improvement, and quality assurance. TQM considers all quality measures taken at all levels and involving all company employees (Kiran, 2016).

The purpose of **COBIT** is to bridge the gap by providing a foundation that is closely linked to business objectives while focusing on IT (IT Governance Institute, 2004). The focus of COBIT is the development of clear policies and good practices for security and control in IT for worldwide endorsement by commercial, governmental and professional organizations. Its primary goal is the development of control objectives primarily from the business objectives and needs perspective. COBIT is designed to be used by three distinct audiences: management, users and auditors.

The **EFQM Excellence Model** provides a common framework and language that facilitates the effective sharing of information between organisations; transcending sectorial; cultural and maturity barriers (Fisher, 2011). The EFQM Excellence Model allows people to understand the cause and effect of relationships between what their organization does, and results which it achieves. The beauty of the EFQM Excellence Model is that it can be applied to any organization, regardless of size, sector or maturity.

Best practices is considered by some as a business buzzword used to describe the process of developing and following a standard way of doing activities that any organization can use or implement to get better results. Implementing best practice in the area of maintenance and reliability can help an organization to improve its performance (Gulati, 2009). There are three main barriers to adoption of a best practice; a lack of knowledge about current best practices, motivation to make changes for their adoption, and knowledge and skill required to do so.

Benchmarking - The objective of benchmarking is to understand and evaluate the current position of a business or organization in relation to the best practice and to identify areas and means of performance improvement (Stapenhurst, 2009). Benchmarking involves looking outside a particular business to examine how others achieve their performance levels, and to understand the processes they use. In this way, benchmarking helps explain the processes behind excellent performance. To be Benchmarking effective, it must become an integral part of an ongoing improvement process, the goal being to abreast of ever-improving best practice.

The **Common Assessment Framework** (CAF) is an easy-to-use, free tool to assist public-sector organizations across Europe in using quality management techniques to improve their performance (Goler, 2009). The model is based on the premise that excellent results in organizational performance, citizens/customers, people and society are achieved through leadership driving strategy and planning, people, partnerships, resources and processes. It looks at the organization from different angles at the same time.

The **PDSA Cycle** (Plan-Do-Study-Act) is a systematic series of steps for gaining valuable learning and knowledge for the continual improvement of a product or process. It is also known as Deming Cycle (Deming, 2016). The cycle begins with the Plan step. This involves identifying a goal or purpose, formulating a theory, defining success metrics and putting a plan into action. These activities are followed by the Do step, in which the components of the plan are implemented, such as making a product. Next comes the Study step, where outcomes are monitored to test the validity of the plan for signs of progress and success, or problems and areas for improvement. The Act step closes the cycle, integrating the learning generated by the entire process, which can be used to adjust the goal, change methods or even reformulate a theory altogether. These four steps are repeated over and over as part of a never-ending cycle of continual improvement.

CorSet Framework is a set of methods and models allowing easier analyses and optimal setting of services, processes and an enterprise management (CorSet Framework, 2017). CorSet is process and service oriented and it is suitable for organizations and companies of all types. It is based on the basic reference models that each organization needs to describe **What** and **How** it does. These are key performance and motivation indicators; a process map, an activity catalogue, a service catalogue, a description of the data and information in the organization, and a function of information system.

3 ANP model for prioritization of the managerial methods

The usability of modern managerial methods is analysed by ANP model, which was quantified based on a questionnaire survey among managers. 18 respondents from the different countries and sectors and on different organizational levels attended questionnaire survey in this pilot study. Those respondents had to answer 7 questions about the usability of the managerial methods – how often they use them, what an issue they solve it with selected methods, but also where they come from, what a position in the work they have.

Consequently, the ANP model was constructed and calculated based on the collected information from questionnaire. The ANP model was chosen because it is more complex multicriteria decision-making method than the AHP. It is a powerful tool to solve the decision-making problem with complex relationships of elements, which have substantial impacts on the decision. The approach of construction was provided as follow: Firstly, two models were created describing feedback systems with two clusters. From those models were found what methods are used either very often or very less and mostly for project issue, and less for other issue. To analyse usability of selected managerial methods in terms of frequency and purpose the aggregate ANP model was prepared and quantified based on the results of a questionnaire survey among managers. We use the ANP model because we want to reflect cyclic dependencies between problem elements described in the following paragraph.

3.1 ANP model structure

The structure of the whole model *Evaluation* in the **Figure 2** describes the whole problem. The model *Evaluation* has four clusters with dependencies between their elements. The first cluster consists of the main goal *Evaluation of The Managerial Methods.* The cluster *Frequency of use* consists of three time elements: *Frequently*, *Sometimes*, and *Rarely*. The next cluster *Purpose of use* has four elements: *Project issue*, *Strategy issue*, *Decision making issue* and *Other issues.* The fourth cluster *The managerial methods* consists of fourteen elements-alternatives of all investigated managerial methods, which have been described above. This last cluster has cyclic dependencies with the previous two. These pairs of clusters represent a feedback system with two components and two partial models *Frequency* (with clusters *Frequency of use* and *The managerial methods*) and *Purpose* (with clusters *Purpose of use* and *The managerial methods*).

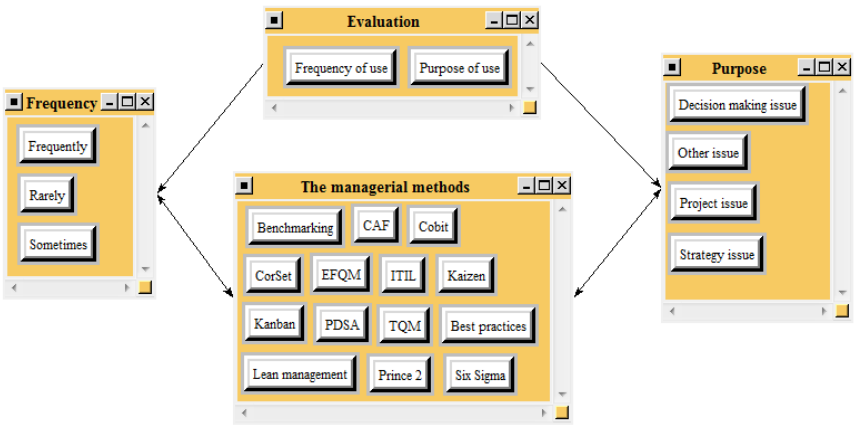


Figure 2 Network structure of the ANP model of the managerial methods problem

The super-matrix of the model *Evaluation* describes the outer and inner dependencies, impact and preferences of the model elements (**Table 2**). The preferences of each supermatrix column were calculated based on questionnaire survey. The consistency test did not need to be done because the preferences were calculated directly from the distribution of the number of individual responses using the scoring method and not from the pairwise comparison method.

	Goal	Frequency	Purpose	Frequently	Sometimes	Rarely	Project issue	Strategy issue	Decision making issue	Other	ITIL	Prince2	Kanban	Kaizen	Lean management	Six Sigma	TQM	Cobit	EFQM Excel. Model	Best practice	Bench marking	CAF	PDSA cycle	CorSet FW
Goal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Frequency	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Purpose	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Frequently	0	0.333	0	0	0	0	0	0	0	0	0.842	0.900	0.684	0.556	0.500	0.600	0.227	0.263	0.158	0.238	0.056	0.176	0.111	0.130
Sometimes	0	0.333	0	0	0	0	0	0	0	0	0.105	0	0.158	0.278	0.333	0.200	0.545	0.474	0.474	0.571	0.222	0	0	0.087
Rarely	0	0.333	0	0	0	0	0	0	0	0	0.053	0.100	0.158	0.167	0.167	0.200	0.227	0.263	0.368	0.190	0.722	0.824	0.889	0.783
Project issue	0	0	0.25	0	0	0	0	0	0	0	0.842	0.368	0.294	0.278	0.474	0.368	0.100	0.059	0.158	0.381	0.150	0	0.118	0.056
Strategy issue	0	0	0.25	0	0	0	0	0	0	0	0.105	0.579	0.471	0.444	0.263	0.263	0.150	0.294	0.211	0.238	0.250	0.211	0.176	0.222
Decision-making issue	0	0	0.25	0	0	0	0	0	0	0	0.053	0.053	0.118	0.167	0.105	0.316	0.500	0.412	0.263	0.333	0.250	0.316	0.118	0
Other	0	0	0.25	0	0	0	0	0	0	0	0	0	0.118	0.111	0.158	0.053	0.250	0.235	0.368	0.048	0.350	0.474	0.588	0.722
ITIL	0	0	0	0.0256	0.0294	0.1321	0.3721	0.0411	0.0141	0.0217	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Prince2	0	0	0	0.0513	0	0.1321	0.0698	0.1644	0.0563	0.0435	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kanban	0	0	0	0.0256	0.0147	0.1415	0	0.1233	0.0986	0.0217	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kaizen	0	0	0	0.0128	0.0735	0.1132	0.0698	0.1233	0.0704	0.0435	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lean management	0	0	0	0.0256	0.1618	0.0472	0.0698	0.0685	0.1127	0.0217	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Six Sigma	0	0	0	0.0385	0.0882	0.0849	0.1163	0.1096	0.0986	0.0435	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TQM	0	0	0	0.0513	0.1176	0.0566	0.0465	0.0548	0.0986	0.0435	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cobit	0	0	0	0.0385	0.1176	0.0660	0.0930	0.0959	0.0845	0.0435	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EFQM Excellence	0	0	0	0.0897	0.1176	0.0283	0.0233	0.0548	0.0282	0.1522	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Best practices	0	0	0	0.1026	0.0441	0.0660	0.0930	0.0822	0.1268	0.0870	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Benchmarking	0	0	0	0.1282	0.0588	0.0377	0.0465	0.0137	0.0704	0.1087	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CAF	0	0	0	0.1282	0.0588	0.0377	0	0.0274	0.0282	0.1739	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PDSA cycle	0	0	0	0.1410	0.0588	0.0283	0	0.0137	0.0845	0.0870	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CorSet Framework	0	0	0	0.1410	0.0588	0.0283	0	0.0274	0.0282	0.1087	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2 Super matrix of the ANP model of the managerial methods problem

The model structure was provided in SuperDecisions software. For its calculations were prepared matrices in MS Excel. The limit super-matrices of both models were obtained as a sequence of the second power of the preceding matrices, the first ones were the weighted supe-matrices. The final values were taken from the 64th power of the weighted supermatrix. The difference between two consecutive values of each synthesized preference is less than 10^{-12}.

3.2 Results

The results of all three ANP models and the final evaluation of the managerial methods are shown in **Table 3** and **Figure 3.** According to the model *Frequency*, the most used methods are PDSA cycle and CorSet Framework but the difference among all methods is very small (from 0.66 to 0.78). The model *Purpose* shows greater difference (from 0.033 to 0.145) and the most used method is ITIL. From the calculation of the whole model *Evaluation*, we have found out that ITIL is the most used method and the CorSet Framework is the less used method.

Importance of managerial methods	**ITIL**	**Prince2**	**Kanban**	**Kaizen**	**Lean management**	**Six Sigma**	**TQM**	**Cobit**	**EFQM Excellence Model**	**Best practices**	**Benchmarking**	**CAF**	**PDSA cycle**	**CorSet Framework**
Evaluation	0.094	0.086	0.080	0.077	0.074	0.071	0.069	0.068	0.067	0.067	0.065	0.062	0.061	0.060
Frequency	0.067	0.070	0.067	0.066	0.067	0.068	0.069	0.068	0.073	0.074	0.077	0.077	0.078	0.078
Purpose	0.142	0.089	0.059	0.080	0.070	0.098	0.059	0.083	0.056	0.096	0.053	0.044	0.037	0.033

Table 3 Results of three ANP models of the managerial methods problem

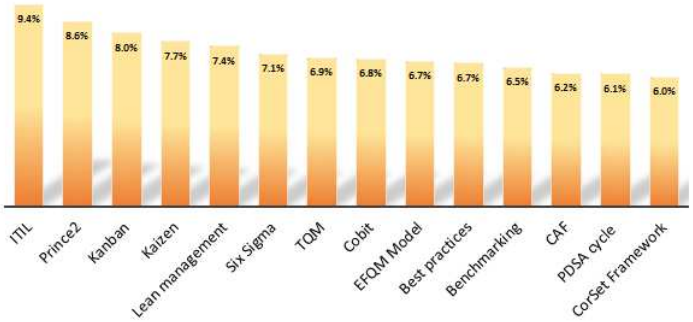


Figure 3 Final importance of the managerial methods

Finding that ITIL is the most commonly used was the kind of prediction because people from IT sector were mainly interviewed. The second most widespread is the Prince2, consequently the Kanban, Kaizen, Lean management, Six Sigma TQM, Cobit, EFQM Model, Best practices, Benchmarking, CAF, PDSA cycle and Corset Framework.

4 Conclusions

In summary, this research attempted to analyse suitability, importance or usability of modern managerial methods using the proper ANP model, which was quantified, based on a questionnaire survey among managers. 18 respondents coming from the different countries and sectors attended pilot questionnaire survey. Those respondents had to answer seven question about the usability of the managerial methods – how often they use them, what an issue they solve it with selected methods, but also where they come from, what a position in the work they have.

Consequently, the ANP model was constructed of the collected information from questionnaire. Firstly, two ANP models *Frequency* and *Purpose* were computed. From those models were found that methods are used either very often or very less and mostly for project issue, and less for other issue.

The whole ANP model *Evaluation* pointed out that the most often used method is ITIL with 9.4 % of importance and the less used method is CorSet Framework with 6 % of importance. The reason why CorSet Framework ended as last is that this is a new method since couple of last years and managers do not know it.

The authors would like to highlight that this paper shows results of small pilot study. The authors will continue in this research based on a large-scale questionnaire survey that will also produce results for individual economic sectors and management levels.

Acknowledgements

This research is supported by the Internal Grant Agency of the University of Life Sciences Prague – project IGA PEF 20171021 - *Model support for choosing the most appropriate management methodology for selected sectors in Czech and Slovak Republic*.

References

[1] Anderson, D. J.: *Kanban: Successful Evolutionary Change for Your Technology Business*. Blue Hole Press, Sequim,Washington, 2010.
[2] Cartlidge, A. et al.: *Úvodní prehled ITIL*. IT SMF, London, UK, 2011. Retrieved Feb. 15, 2017, from http://itsmf.cz/wp-content/ uploads/ 2013/12/Uvodni_prehled_ITIL_2011.pdf.
[3] CorSet Framework Ltd. *CorSet Framework*. London, UK, 2017. Retrieved Feb. 13, 2017, from http://www.corsetframework.eu/index.php/en/corset-framework.
[4] Deming, W. E.: *PDSA Cycle*. The W. Edwards Deming Institute. Ketchum, USA, 2016. Retrieved Feb. 4, 2017, from https://deming.org/management-system/pdsacycle.
[5] Fisher, M.: *Introducing the EFQM Excellence Model*. Amazon Kindle Direct Publishing, 2011.
[6] Goler, B.: *The Common Assessment Framework Handbook*. Piggott Black Bear Limited, Cambridge, 2009.
[7] Gulati, R.: *Maintenance Best Practices*. Industrial Press Inc., New York, 2nd edition, 2009.
[8] Hinde, D.: *PRINCE2 Study Guide.* TJ International, Padstow, UK, 2009.
[9] Imai, M.: *Kaizen: Metoda, jak zavést úspornejší a flexibilnejší výrobu v podniku*. Computer Press, Brno, 2008.
[10] IT Governance Institute: *CoBiT in Academia*. Rolling Meadows, USA, 2004. Retrieved Feb. 12, 2017, from http://www4.gvsu.edu/CANNOND/Cobit_Student_Book.pdf.
[11] Kiran, D.R.: *Total Quality Management – Key Concepts and Case Studies.* Butterworth-Heinemann, United Kingdom, 1st edition, 2016.
[12] OGC: *Introduction to the ITIL service lifecycle.* The Stationery Office, United Kingdom, 2010.
[13] Saaty, T.L.: *Decision Making For Leaders: The Analytic Hierarchy Process*. University of Pittsburgh: RWS Publications. 3rd edition, 2012.
[14] Saaty, T.L.: *Theory and Applications of Analytic Network Process: Decision Making with Benefits, Opportunities, Costs and Risks.* University of Pittsburgh: RWS Publications, 2009.
[15] Stapenhurst, T.: *The Benchmarking Book: A How-to-Guide to Best Practice for Managers and Practitioners.* Butterworth-Heinemann, United Kingdom, 2009.
[16] Töpfer, A.: *Six Sigma: koncepce a príklady pro rízení bez chyb.* Computer Press, Brno, 1st edition, 2008.

The effect of the deaths from chronic ischemic heart disease versus acute coronary syndrome on life expectancy among the Slovak population

Beata Gavurova [1], Tatiana Vagasova[2]

Abstract. In recent years, high mortality from chronic ischemic heart disease and acute coronary syndrome have burdened economic as well as health system of the Slovak Republic considerably. By eliminating these deaths, the life expectancy could be prolonged what is in the interest of each country in the world. On the basis of the method of constructing the abridged life tables, the potential gains in life expectancy are calculated. They are computed from the mortality reports (1996-2014) for both diseases by sex. Moreover, the added years, that would be gained by eliminating causes of deaths, are decomposed by five-year age groups. This method is not influenced by the age and size population structure, what offers a clear advantage over the other techniques, such as the years of potential life lost calculation. Our findings pose immediate challenges to policy makers and raise a discussion about better health interventions needed for the reduction in burden of diseases. The targeted primary prevention programs result in the better health status of population, future increasing the life expectancy as well as higher labour productivity and economic growth.

Keywords: potential gains in life expectancy, the method of constructing the abridged life tables, chronic ischemic heart disease, acute coronary syndrome.

JEL Classification: C44, I10
AMS Classification: 91B32

1 Introduction

In epidemiological research, measurements of disease burden are associated with a threat to population that shorten years of life. Burden of diseases is usually expressed by indicators of mortality and morbidity. According to Pol and Thomas [14], diseases are typically categorised as either acute characterising by short duration, rapid onset, and usually death, or chronic defining by lengthy progression, slow onset, and long duration of care. In a current modern society, the leading causes of death are chronic conditions such as cancer, diabetes, chronic respiratory diseases, but mostly cardiovascular diseases as stated in [4][5][18].

The World Health Organisation [17] estimates that the number of deaths caused by these diseases is about 4 million per year in Europe. Kesteloot et al. [7] show that cardiovascular diseases mortality rates have been falling since 1970 in Western Europe, but in Eastern Europe remain comparatively high as in [6]. Specifically, Nichols et al. [13] found that cardiovascular diseases mortality rates decreased by 17-18% in Slovakia between 2001 and 2010, however, they decreased by 32-33% in France and by 31-34% in Germany during approximately the same time period.

It raises a question how many additional years of life expectancy would be gained, if the main causes of deaths were eliminated. We focus on the comparison between acute and chronic heart disease, specifically acute coronary syndrome and chronic ischemic heart disease. We would like to find out which of them shorten a life of population in a larger extent.

This issue can be measured, on one hand, by general measures such as Quality Adjusted Life Years (QALY) and Disability Adjusted Life Years (DALY), on the other hand, by more specific methods such as Potential Years of Life Lost (PYLL) and Potential Gain in Life Expectancy (PGLE) as per [8][12][18]. Lai and Hardy [8] compared using PYLL with PGLE. While PYLL expresses the impact only premature deaths on years of life lost when the age limit of 75 years is frequently chosen, PGLE can reflect the loss of life expectancy caused by a certain disease and provide a numerical indicator of survival if the cause of death is eliminated. Moreover, PGLE is not influenced by the age structure of population, permitting comparisons between diseases. PYLL gives

[1] Technical University of Košice, Faculty of Economics, Němcovej 32, Košice, Slovakia, e-mail: beata.gavurova@tuke.sk.
[2] Technical University of Košice, Faculty of Economics, Němcovej 32, Košice, Slovakia, e-mail: tatiana.vagasova@tuke.sk.

greater weight to deaths at a younger age and lower weight to deaths at older age as per [3]. Generally, there are many other studies that concerned with quantifying the effect of elimination of diseases deaths on life expectancy by PGLE, for example, Conti et al. [2], Liu et al. [9], Mackenbach et al. [10], Tsai et al. [15], Wang et al. [16], etc.

These techniques can be used to distinguish competing risks of diseases and to identify a larger threat to the population's health. They are also required in planning health care needs in the future, such as the long-term health services, or in predicting national priorities in prevention strategies.

The aim of this paper is to calculate and compare the impact of deaths from chronic ischemic heart disease and acute coronary syndrome on life expectancy of the Slovak population by sex from 1996 to 2014 and to find out the most threatened age groups of population.

2 Data and methods

Under the conditions of the contract, data on the number of deaths by five-year age groups for the period 1996-2014 were obtained from National Health Information Centre of Slovakia. Data on the mid-year population at the age groups in every year were downloaded from the Statistical Office of the Slovak Republic.

PGLE reflects how many years on average a person would still live, if a given cause of death was eliminated. In other words, PGLE expresses years of life lost resulting from a certain disease in an age group. So, life expectancy could be extended for these years. The higher PGLE, the higher impact of the disease on life expectancy is.

For calculation of PGLE it was needed to construct the life tables and the specific life tables regarding causes of deaths in the Slovak Republic in every year and an age group separately by the methodical tutorials of Demographic Research Centre [11] and National Vital Statistics Reports [1]. We examined the life expectancy (e_x) expressing the all causes of deaths and cause-eliminated life expectancy ($e_x^{(-i)}$) by elimination of a certain cause of death. The Tenth Revision of the International Classification of Disease (ICD-10) was used to specify causes of death included in this analysis: chronic ischemic heart disease (I25) and acute coronary syndrome (I20-I22).

Cause-eliminated life expectancy ($e_x^{(-i)}$) was the result from analysis of deaths caused by a certain disease and was calculated from the abridged life table minus causes of death. According to the Demographic Research Centre [11], the first step is the calculation of the probabilities of survival ($_np_x$) from the all-caused abridged life tables with the formula:

$$_n p_x = 1 - {}_n q_x \tag{1}$$

where x – the exact age; n – the number of years in the age interval; $_nq_x$ - the probability of dying between the beginning of an age interval and before reaching the end of that age interval.

Then, the probabilities of death eliminating the i_{th} cause ($_nq_x^{(-i)}$) were estimated by:

$$_n q_x^{(-i)} = 1 - {}_n p_x^{\left(\frac{{}_nD_x - {}_nD_x^i}{{}_nD_x}\right)} \tag{2}$$

where $_nD_x$ - the number of deaths in the age interval x to $x + n$ for all causes; $_nD_x^i$ - the number of deaths in the age interval x to $x + n$ attributable to the i_{th} cause of death.

Arias et al. [1] report the number of person-years lived ($_nL_x^{(-i)}$) in the age interval x to $x + n$ estimated for ages 0, 1, 5, 10,……, 95 by the formula:

$$_n L_x^{(-i)} = (n - {}_n f_x) \cdot l_x^{(-i)} + {}_n l_x \cdot l_{x+n}^{(-i)} \tag{3}$$

where n=1 for x=0, n=4 for x=1, and n=5 for x=5, 10, ..., 95; $_nl_x$ - the number of persons from the original life table who survive to the beginning of each age interval; $l_x^{(-i)}$ - the number of survivals from life table due to the i_{th} causes; L_x – the number of person-years from the original life table within an age interval x to $x + n$, and the quantities $_nf_x$ were estimated from the all-cause life table by:

$$_n f_x = \frac{n \cdot {}_n l_x - {}_n L_x}{l_x - l_{x+n}} \tag{4}$$

The last step is to calculate the number of person-years lived after exact age x ($T_x^{(-i)}$) by:

$$T_x^{(-i)} = L_x^{(-i)} + L_{x+1}^{(-i)} + \cdots + L_{95+}^{(-i)} \tag{5}$$

Finally, the cause-eliminated life expectancy ($e_x^{(-ij)}$) is calculated as:

$$e_x^{(-i)} = \frac{T_x^{(-i)}}{l_x^{(-i)}} \tag{6}$$

Subsequently, the PGLE of a disease in a certain year is calculated as the difference between cause-eliminated life expectancy ($e_x^{(-ij)}$) and life expectancy (e_x) in the same year.

$$PGLE = e_x^{(-i)} - e_x \tag{7}$$

3 Results

This section examines the progress of the impact of deaths from specific causes on life expectancy by sex in Slovakia during 1996-2014. Moreover, it compares the current burden of diseases among the age groups of population by sex.

3.1 Trends of potential gains in life expectancy at birth

Figure 1 depicts that potential gains in life expectancy (PGLEs) by elimination of deaths from chronic ischemic heart disease are higher than acute coronary syndrome for both sexes throughout the whole time span. While the chronic ischemic heart disease burdens on females in larger extent than on males, an opposite effect is observed in the acute coronary syndrome, in which males are more burdened compared with females.

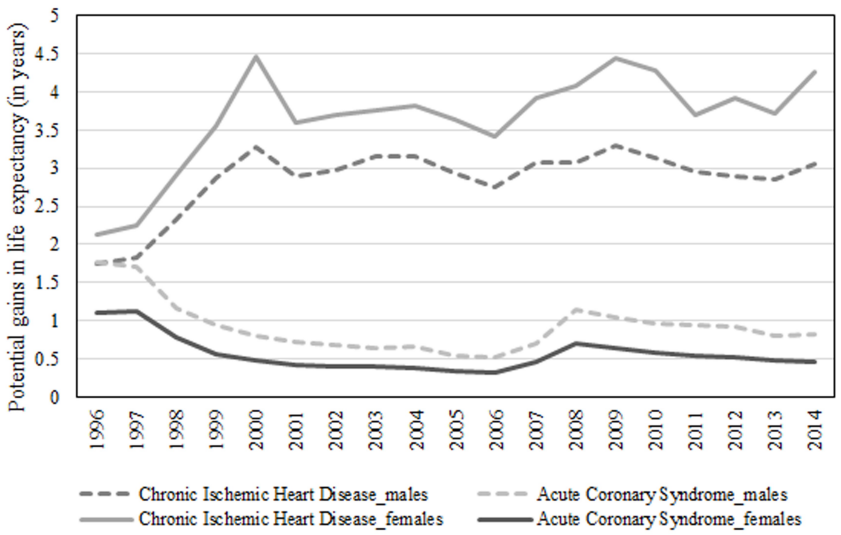


Figure 1 Trends of potential gains in life expectancy at birth (in years) for the Slovak population by elimination of the deaths from chronic ischemic heart disease and acute coronary syndrome by sex, 1996-2014

In 1996, the PGLEs reached the values of 1.75 and 1.77 years for chronic ischemic heart disease and acute coronary syndrome, respectively, among males. Considering females, the PGLE for chronic ischemic heart disease reached a value of 2.13 years compared to 1.1 added years for acute coronary syndrome. The scissors-like differences between these diseases for both sexes have still extended. In 2014, among males, an increase of 3.05 years was calculated for chronic ischemic heart disease (74.3%). On the contrary, the effect of eliminating deaths from acute coronary syndrome decreased on 0.83 years (-53.3%). Similarly, among females, the PGLEs increased on 4.26 years in 2014 (100%) for chronic ischemic heart disease compared with a decrease on a value of 0.46 years for acute coronary syndrome (-57.9%). As one could notice, with the elimination of chronic ischemic heart disease, both sexes experienced growing trend, and falling trend with the elimination of acute coronary syndrome. It means that the burden of chronic ischemic heart disease to population's health is worsening, while the burden of acute coronary syndrome is improving.

3.2 Potential gains in life expectancy at five-year age groups in 2014

The potential added years of life at birth as well as at the other age groups through the elimination in mortality from chronic ischemic heart disease and acute coronary syndrome according to the sex in 2014 are showed in Table 1.

We have mentioned above that the highest impact on life expectancy (4.26 years) was examined after elimination of the chronic ischemic heart disease deaths for females. It means that a woman at birth could be expected to live 4.26 years longer than the actual life expectancy at birth for females (80.39 years). While a man at birth could be expected to live 3.05 years longer than the actual life expectancy at birth for males (73.28 years).

Commonly, as one would appreciate, the elimination of a disease should result in rather lower and lower rise in life expectancy at older ages. However, we can see that the PGLEs of chronic ischemic heart disease mortality elimination continuously increase until age 95+ for both sexes. These results show a strengthening impact of chronic ischemic heart disease on life expectancy, mostly in older people.

	Males					Females				
Age groups	LE without elimination e_x	CIHD $e_x^{(-i)}$	CIHD gain	ACS $e_x^{(-i)}$	ACS gain	LE without elimination e_x	CIHD $e_x^{(-i)}$	CIHD gain	ACS $e_x^{(-i)}$	ACS gain
00-00	73.28	76.33	3.05	74.11	0.83	80.39	84.65	4.26	80.86	0.46
01-04	72.72	75.79	3.07	73.55	0.83	79.84	84.11	4.28	80.30	0.46
05-09	68.80	71.88	3.08	69.64	0.83	75.91	80.19	4.28	76.37	0.47
10-14	63.86	66.94	3.08	64.69	0.83	70.94	75.22	4.28	71.40	0.47
15-19	58.91	61.99	3.08	59.75	0.83	65.96	70.24	4.28	66.42	0.47
20-24	54.05	57.14	3.09	54.88	0.84	61.04	65.33	4.29	61.51	0.47
25-29	49.24	52.34	3.10	50.07	0.84	56.11	60.41	4.29	56.58	0.47
30-34	44.43	47.54	3.11	45.27	0.84	51.18	55.47	4.30	51.64	0.47
35-39	39.66	42.78	3.12	40.50	0.84	46.28	50.59	4.31	46.75	0.47
40-44	34.97	38.11	3.14	35.80	0.84	41.45	45.77	4.32	41.92	0.47
45-49	30.41	33.57	3.16	31.23	0.82	36.65	40.99	4.34	37.11	0.46
50-54	26.08	29.27	3.19	26.87	0.79	31.99	36.35	4.36	32.45	0.46
55-59	22.03	25.26	3.23	22.78	0.75	27.43	31.83	4.40	27.88	0.45
60-64	18.38	21.65	3.27	19.05	0.66	23.08	27.54	4.46	23.52	0.44
65-69	15.07	18.40	3.33	15.65	0.58	18.94	23.50	4.56	19.36	0.41
70-74	12.01	15.43	3.42	12.49	0.48	14.97	19.63	4.66	15.35	0.38
75-79	9.25	12.83	3.58	9.65	0.40	11.35	16.08	4.73	11.67	0.32
80-84	6.93	10.71	3.78	7.25	0.32	8.26	13.07	4.81	8.53	0.26
85-89	5.18	9.29	4.11	5.42	0.25	5.98	10.87	4.89	6.19	0.21
90-94	4.19	8.66	4.47	4.44	0.25	4.49	9.52	5.03	4.67	0.18
95+	4.90	9.68	4.78	5.25	0.35	4.70	9.85	5.15	4.83	0.13

Table 1 Potential gains in life expectancy (in years) at age groups by elimination of the deaths from chronic ischemic heart disease (CIHD), acute coronary syndrome (ACS) by sex, based on 2014 mortality data

As for acute coronary syndrome among both sexes, the PGLEs increase up to the 40-44 years old. According to the hypothesis of the elimination of acute coronary syndrome, the PGLE at the value 0.84 years means the years of life lost which are the same for 20-24 age group up to the 40-44 years old men. This shows that people in working age groups could gain most benefits from the disease elimination, hence, they are most threatened with this disease compared to other age groups of people. The PGLEs improve from 45-49 years old to the end of life. For instance, life expectancy is extended by an additional 0.82 years at ages 45-49, 0.48 years at ages 70-74, etc., approaching zero. Similarly, trend of PGLEs in acute coronary syndrome among women is decreasing from 45-49 age group. Equally, the most threatened age groups among females are working groups, even children. A person at 05-09 age group could be expected to live 0.47 years longer than the actual LE at birth (75.91 years). Evenly, a person at 40-44 aged could live 0.47 years longer than actual LE at this age group (41.45

years). Thus, we found out that people at a younger age are more threatened by acute coronary syndrome, while people at an older age by chronic ischemic heart disease.

4 Conclusion

The aim of this analysis was to quantify the impact of deaths from chronic ischemic heart disease and acute coronary syndrome on life expectancy of the Slovak population by sex from 1996 to 2014 and to reveal the differences between age groups of population. Our results demonstrate competing risks of death as candidate diseases for launching, monitoring and assessing prevention programs in the Slovak health system.

In terms of gained years of life, the highest benefit would be obtained by prevention programs for reduction mortality of chronic ischemic heart disease. They should be primarily aimed at women, mostly at older age groups. Unfortunately, a national program of cardiovascular disease prevention as well as treatment guidelines do not exist in Slovakia. There are any specific prevention or treatment procedures that would take into account gender differences in Slovakia. Henceforth, the health policy makers should focus on eliminating risk factors (e.g. LDL cholesterol, dietary habits, physical activity, smoking) and compensation of already existing diseases (hypertension, diabetes mellitus – type 2, metabolic syndrome, etc.) closely related to the heart diseases.

We have to realise that the extension of life expectancy in Slovakia supports the increased incidence of many chronic diseases that are typical of older population. This issue results in an urgent need to ensure sufficient and available long-term health care that naturally occurring with the ageing population.

On the contrary, deaths from acute coronary syndrome are slightly decreasing for both sexes, although their elimination would have a higher influence on life expectancy for males than females. Since, in Slovakia there is not active health interventions that could monitor and evaluate health behaviour of population, it is difficult to explain why the burden of acute coronary syndrome declines or is higher for men. At least, we can suppose that risk factors such as smoking, alcohol consumption, unhealthy eating, rare general practitioner utilisation for preventive examination, life stress are more typical for males compared to females.

Our results are supported by the fact stated in [14] that acute diseases are more common in developing countries and to younger population, while Slovakia belongs to the developed countries with ageing population in which chronic conditions have a predominance.

Our analysis has its strengths and limitations. We calculated the potential gains in life expectancy what seems to be a better indicator of the burden of disease, unlike other burden of disease indicators, expresses the maximal number of years to which the life expectancy could be prolonged after elimination of a cause of death. Although, total elimination of a certain cause of death is not likely, this indicator provides the real strength to the other competing risks of causes of death. Because of data unavailability of ethnicity and socioeconomic characteristics, we were not able to calculate the potential gains in life expectancy for a specific group of people, thus large variations within population may be concealed.

However, the findings of this analysis can contribute to more effective allocation of scare resources in the Slovak health system, mainly for prevention programs. All actors, who form acute or chronic diseases prevention, such as patients, primary health care, hospitals, media, schools, etc., play an important role in achieving prolonged life expectancy.

Acknowledgements

The work was supported by the VEGA Project No. 1/0945/17 ”Research of economic quantification of marketing processes aimed at increasing the value for the patient, multidimensional analyses of the marketing mix of medical facilities, and quantifying their importance in the process of the system creation for measurement the quality and effectiveness of the Slovak health system”. Our thanks go to the National Health Information Center of Slovakia for providing the Slovak mortality database as well as for a long-term support of our research activities within health policy.

References

[1] Arias, E., Heron, M., and Tejada-Vera, B.: *United States Life Tables Eliminating Certain Causes of Death, 1999–2001. National vital statistics reports*. National Center for Health Statistics, USA, 2013.

[2] Conti, S., Farchi, G., Masocco, M., Toccaceli, V., and Vichi, M.: The impact of the major causes of death on life expectancy in Italy. *International Journal of Epidemiology* **28** (1999), 905-910.

[3] Gardner, J. W., and Sanborn, J. S.: Years of potential life lost (YPLL) - what does it measure*? Epidemiology* **1** (1990), 322-329.

[4] Gavurová, B., Kováč, V., and Vagašová, T.: Standardised mortality rate for cerebrovascular diseases in the Slovak Republic from 1996 to 2013 in the context of income inequalities and its international comparison. *Health Economics Review* **7** (2017), 1-12.
[5] Gavurová, B., and Vagašová, T.: Regional differences of standardised mortality rates for ischemic heart diseases in the Slovak Republic for the period 1996–2013 in the context of income inequality. *Health Economics Review* **6** (2016), 1-12.
[6] Graham, I., Atar, D., and Borch-Johnsen, K. et al.: European guidelines on cardiovascular disease prevention in clinical practice: full text. Fourth Joint Task Force of the European Society of Cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). *European Journal of Cardiovascular Prevention and Rehabilitation* **14** Suppl 2 (2007), S1-113.
[7] Kesteloot, H., Sans, S., and Kromhout, D.: Dynamics of cardiovascular and all-cause mortality in Western and Eastern Europe between 1970 and 2000. *European Heart Journal* **27** (2006), 107-113.
[8] Lai, D., and Hardy, R. J.: Potential gains in life expectancy or years of potential life lost: impact of competing risks of death. *International Journal of Epidemiology* **28** (1999), 894-898.
[9] Liu, P., Li, Ch., and Wang, Y. et al.: The impact of the major causes of death on life expectancy in China: a 60-year longitudinal study. *BMC Public Health* **14** (2014), 1193.
[10] Mackenbach, J. P., Kunst, A. E., and Lautenbach, H. et al.: Gains in life expectancy after elimination of major causes of death: revised estimates taking into account the effect of competing causes. *Journal of Epidemiology and Community Health* **53** (1999), 32–37.
[11] Mészáros, J.: Methodological guide. *The calculation of life tables. The calculation of Potential Years of Life Lost.* INFOSTAT Demographic Research Centre, Bratislava, 2000.
[12] Murray, C. J. L., and Lopez, A. D.: *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Global Burden of disease and Injury Series*. Harvard University Press, Cambridge, 1996.
[13] Nichols, M., Townsend, N., Scarborough, P., and Rayner, M.: Cardiovascular disease in Europe 2014: epidemiological update. *European Heart Journal* **35** (2014), 2950-2959.
[14] Pol, L. G., and Thomas, R. K.: *The Demography of Health and Healthcare*. Springer Science+Business Media, Dordrecht, 2013.
[15] Tsai, S. P., Lee, E. S., and Hardy, R. J.: The Effect of a Reduction in Leading Causes of Death: Potential Gains in Life Expectancy. *American Journal of Public Health* **68** (1978), 966-971.
[16] Wang, G. D., Lai, D. J., Burau, K. D., and Du, X. L.: Potential gains in life expectancy from reducing heart disease, cancer, Alzheimer's disease, kidney disease or HIV/AIDS as major causes of death in the USA. *Public Health* **127** (2013), 348-356.
[17] World Health Organisation: *Deaths from cardiovascular diseases*. Global Health Observatory, Geneva, 2014. Available at: http://gamapserver.who.int/gho/interactive_charts/ncd/mortality/cvd/atlas.html.
[18] World Health Organisation: *WHO methods and data sources for global burden of disease estimates 2000-2015*. Department of Information, Evidence and Research, Geneva, 2017.

Transportation problem with degressively stepped costs

Vojtěch Graf[1], Dušan Teichmann[2], Michal Dorda[3]

Abstract. The transportation problem belongs to one of elementary decision problems we can meet in practice. The problem can be solved with linear programming methods. In specialized literature, we can find many possible modifications of the transportation problem which have been formulated and solved on the basis of many real situations. This fact reveals its broad area of possible applicability. The article deals with a modification of the transportation problem; the modification consists in a specific nonlinear form of the objective function which has not been found in literature. In the article, it is considered that a carrier offers a lower carriage to a customer if a certain (minimum) number of goods is transported. Such type of the transportation problem has already been considered in the past but the objective criterion has been assumed to be continuous nonlinear. In our modification nonlinearity of the objective criterion is discontinuous (stepped) and is decreasing (the carriage per unit is decreasing with the increasing number of transported goods). The task can be solved as balanced (the total supply and the total demand are equal) or unbalanced (the total supply and the total demand are not equal). A linear mixed integer mathematical model which models the modification of the transportation problem is presented in the article. To test the model and to find a dependence of the calculation time on input data some calculation experiments are presented in the article. The calculation experiments are carried out using optimisation software Xpress – IVE.
Keywords: transportation problem, nonlinear, linear programming

JEL Classification: C61
AMS Classification: 90C15

1 Introduction and state of the art

The transportation problem including its possible modifications is discussed by many authors. The original transportation problem was defined by L. Hithcook as a linear programming model in publication [1]. Many authors followed in research on the transportation problem – for example G. B. Dantzig applied the simplex method to solve the transportation problem [2]. In addition, G. B. Dantzig also created a special algorithm for solving the transportation problem. Mathematical models of the basic transportation problems are well-known; readers can found them almost in each publication which is devoted to operations research and its applications.

As mentioned earlier in the text, many modifications of the basic transportation problems were defined in the past. In paper [3] a mathematical model of the transportation problem with a different discontinuous piecewise linear cost function is presented as well. However, the model is solved using a genetic algorithm. In publication [4] a mathematical model of a multi-stage transportation problem or a mathematical model of a min-max transportation problem are presented. For the multi-stage transportation problem it is typical that transportation of goods between the sources and destinations is realised for example via logistics terminals. The min-max transportation problem assigns the numbers of goods transported between the sources and destinations as evenly as possible. Some interesting modifications of the transportation problem are discussed in [5], [6] a [7]. In publication [5] a modification of the transportation problem with fixed charges is presented; if any non-zero number of goods is transported between the source and the destination, the fixed fee is charged. Publication [6] presents some modifications of the transportation problems with sanctions. The transportation problem with the sanctions is always unbalanced, the sanctions are imposed if any source is not depleted or any destination is not satisfied. In publication [7] some variants of the transportation problem with nonlinear objective functions are discussed. All the nonlinear functions are increasing; the models that are presented in the publication have the objective function increasing progressively or degressively. The authors of the publication transform the nonlinear objective functions into functions that are piecewise linear – see an example in Fig. 1, where C_{KL} are the total transport costs and X_{KL} is the number of transported goods.

[1] VŠB-TU Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17.listopadu 15, Ostrava, 70833, vojtech.graf@vsb.cz
[2] VŠB-TU Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17.listopadu 15, Ostrava, 70833, dusan.teichmann@vsb.cz
[3] VŠB-TU Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17.listopadu 15, Ostrava, 70833, michal.dorda@vsb.cz

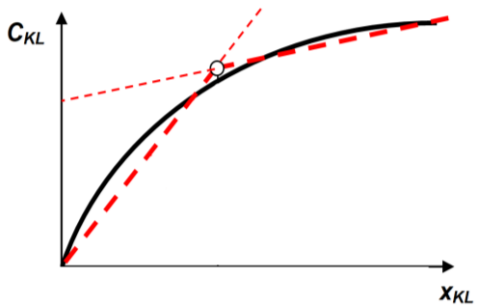


Figure 1 The cost function for decreasing unit transportation costs [7]

To plan how the given destinations (or the customers) should be supplied from the given sources (or the store-houses), a mathematical model of the well-known transportation problem can be employed. However, the basic transportation problem model is very simple and it cannot often reflect real distribution problems. Therefore, it is necessary to continue in improving the transportation problem models; such improvements usually lead to adding additional constraints to model real distribution systems.

In most transportation problem models, it is considered that the costs resulting from distribution of goods should be as minimal as possible from the point of view of carriers which operate the modelled distribution system. But in practice, it often happens that users of the distribution system submit their orders in which the number of goods transported from the sources to the destinations is already defined on the basis of price quotations of the individual carriers. In addition, in the real distribution systems it is usual that the unit transportation costs decrease with the increasing number of goods that is transported. In such cases, the unit transportation costs do not decrease continuously, but they are decreasing step-wisely – an example of such gradual changes in the transportation costs is depicted in Fig. 2. The individual decreases in the unit transportation costs can be represented by quantity discounts that are offered by the carrier. Figure 2 shows the decreases of the unit transportation costs using so called coefficients of quantity discount. The basic unit transportation costs are then multiplied with the corresponding coefficient depending on the number of transported goods for all the possible directions of transport – from each source to each destination. We can see in the figure that if the number of transported goods is less than 5, then no quantity discount is offered. If 5 up to 9 units of goods are transported then the coefficient of quantity discount is equal to 0.7. That means the unit transportation costs applied in this case equal to 70% of the basic transportation costs (with no quantity discount). If at least 10 units of goods are transported then the coefficient of quantity discount is 0.4 and, therefore, the unit transportation costs equal to 40% of the unit transportation costs without quantity discount.

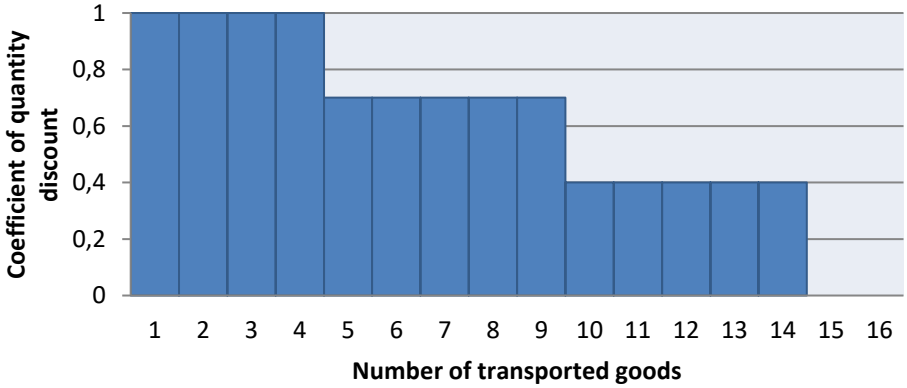


Figure 2 The influence of the number of transported goods on the unit transportation costs

Let us define a new term – a price zone. The price zone corresponds to an interval for which it holds that the same coefficient of quantity discount is applied. Figure 1 shows the case for which three price zones are defined. In general, the number of the price zones can be different for the individual directions of transport (from the source to the destination).

Our goal is to create a mathematical model which optimises the costs from the point of view of the distribution system users. That means the users of the distribution system define the numbers of goods transported between

the individual sources and destinations so that the total costs of transportation associated with supply of the destinations are as minimal as possible.

2 Problem formulation

Let us consider a distribution system which supplies some processing plants (the destinations) from some mines (the sources) with raw material (goods). Let a set *I* of the sources and a set *J* of the destinations be given. For each source $i \in I$ its capacity a_i must be known, for each destination $j \in J$ its demand b_j must be given. A set *K* of the price zones must be also defined, using the price zones different transportation costs are modelled as discussed in the previous section. Please note that in the presented mathematical model it is assumed that all the directions of transportation (from the individual sources to the destinations) have the same number *p* of the price zones. For each direction of transportation – from each source $i \in I$ to each destination $j \in J$ – the basic transportation costs denoted c_{ij} are defined, the basic costs correspond to the unit transportation costs for the first price zone (the coefficient of quantity discount is equal to 1). For each price zone $k \in K$ in the direction from the source $i \in I$ to the destination $j \in J$ an upper number Q_{ijk} of transported goods is defined and for each price zone $k \in K \setminus \{1\}$ a lower number Q_{ijk-1} of transported goods is given. For the first price zone in the direction from the source $i \in I$ to the destination $j \in J$ it holds that the number of transported goods lies in $\langle 0; Q_{ij1} \rangle$, for the second price zone in $\langle Q_{ij1}; Q_{ij2} \rangle$ and for the last price zone in $\langle Q_{ijp-1}; Q_{ijp} \rangle$, where $Q_{ijp} = \min\{a_i, b_j\}$ and $|K| = p$. It is clear that the individual price zones are mutually subsequent. For each price zone $k \in K$ and the direction from the source $i \in I$ to the destination $j \in J$ a coefficient of price discount $r_{ijk} \in \langle 0;1 \rangle$ is given; it holds that $r_{ij1} > r_{ij2} > ... > r_{ijp}$. Our goal is to plan how goods should be transported from the individual sources to the individual destinations so that the demands of all the destinations are satisfied and the total transportation costs are minimal.

3 Mathematical model

To create the mathematical model, two groups of variables must be defined. The first group of the variables models how many units of goods are transported in the corresponding price zone. The variables are denoted x_{ijk} – the values of the variable express how many units of goods are transported from the source $i \in I$ to the destination $j \in J$ within the price zone $k \in K$. The second group of the variables corresponds to the auxiliary binary variables denoted u_{ijk}. If $u_{ijk} = 1$, then the costs of transportation from the source $i \in I$ to the destination $j \in J$ belong to the price zone $k \in K$. If $u_{ijk} = 0$, then the costs of transportation from the source $i \in I$ to the destination $j \in J$ do not fall into the price zone $k \in K$ (that means the price discount that corresponds to the price zone is not applied).The presented mathematical model does not consider a capacity of the vehicles transporting goods from the individual sources to the destinations. It is assumed that the capacity of the vehicles is sufficient. And finally, the unbalanced transportation problem with the total supply in the sources which is greater than the total demand of the destinations is assumed; the modification occurs most often in practice.

The mathematical model can be defined in the following form:

$$\min f(x,u) = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ij} \cdot r_{ijk} \cdot x_{ijk} \tag{1}$$

subject to:

$$\sum_{j \in J} \sum_{k \in K} x_{ijk} \le a_i \text{ for } i \in I, \tag{2}$$

$$\sum_{i \in I} \sum_{k \in K} x_{ijk} = b_j \text{ for } j \in J, \tag{3}$$

$$x_{ijk} \ge Q_{ijk-1} \cdot u_{ijk} \text{ for } i \in I, j \in J, k \in K \setminus \{1\}, \tag{4}$$

$$x_{ijk} \le Q_{ijk} \cdot u_{ijk} \text{ for } i \in I, j \in J, k \in K, \tag{5}$$

$$\sum_{k \in K} u_{ijk} \le 1 \text{ for } i \in I, j \in J, \tag{6}$$

$$u_{ij1} \le x_{ij1} \text{ for } i \in I, j \in J, \tag{7}$$

$$x_{ijk} \geq 0 \text{ for } i \in I, j \in J, k \in K, \tag{8}$$

$$u_{ijk} \in \{0,1\} \text{ for } i \in I, j \in J, k \in K. \tag{9}$$

Formula (1) represents the objective criterion of which values express the total costs of transportation between the sources and the destinations. The group of constraints (2) ensures that no capacity of any source is exceeded (we cannot transport from any source the number of goods that is greater than its capacity). The constraints cover two possible situations that may occur – they correspond to the balanced transportation problem or the unbalanced transportation problem with the total supply greater than the total demand. The group of constraints (3) models that the demand of each destination $j \in J$ must be satisfied. The groups of constraints (4) and (5) define the price zones and their width. The group of constraints (6) ensures that for each direction of transportation at most one price zone is applied. The group of constraints (7) prevents the binary variables u_{ij1} from taking value 1 if $x_{ij1} = 0$. Constraints (8) and (9) define the domains of definition for the variables used in the model. It results from the problem definition that the numbers of the transported goods are integer in practice. Therefore, constraints (8) can be defined alternatively – see constraints (10) that ensure the numbers of goods transported between the individual sources to the individual destinations are non-negative integers (Z_0^+):

$$x_{ijk} \in Z_0^+ \text{ for } i \in I, j \in J, k \in K. \tag{10}$$

4 Experiments with the model and their results

To test the presented mathematical model and to find out how long calculation times are needed for solving the model, several calculation experiments were carried out. All the optimisation experiments were realised using optimisation software Xpress-IVE which employs branch and bound methods for solving such tasks. The individual experiments differ in the number of the sources, destinations and price zones and in the domain of definition of the variables x_{ijk}. For each experiment, it holds that $m = n = p$. Therefore, in the following text we denote $m = n = p = S$.

Obtained results of the experiments are compared with one another. Note that a personal computer with a 3.3 GHz processor AMD FX – 8300 Eight-Core, 8 GB of RAM and a full Xpress-IVE license was employed.

Based on the experiments we can conclude that the presented mathematical model is functional. After finishing each experiment we recorded the calculation time needed for solving the model. The calculation times are expressed in seconds and were obtained for the non-negative integer and non-negative variables x_{ijk} separately – the calculation times are summarised in Table 1. Notation R_0^+ is used for the nonnegative values of x_{ijk} and Z_0^+ for the non-negative integer values.

S	The number of variables x_{ijk}	The number of variables u_{ijk}	The number of the constraints excluding constraints (8) and (9)	The calculation times for $x_{ijk} \in R_0^+$ [s]	The calculation times for $x_{ijk} \in Z_0^+$ [s]	The differences between the calculation times [s]
3	9	9	123	0.1	0.1	0
5	125	125	535	0.1	0.1	0
10	1 000	1 000	4 120	0.2	0.2	0
15	3 375	3 375	13 755	0.8	0.4	0.4
20	8 000	8 000	32 440	11.9	7.9	4.0
25	15 625	15 625	63 175	38.3	20.5	17.8
30	27 000	27 000	108 960	123.2	95.7	27.5

Table 1 Results of the experiments

In total, 14 optimisation experiments were carried out including the experiments that were realised to test the model. 7 experiments were carried out on the assumption that the variables x_{ijk} can take the nonnegative values and for the remaining experiments it was assumed that the variables x_{ijk} can be only the nonnegative integers.

We can say based on the experiments that an increase of the calculation times is remarkable for the models if the cardinality of sets I (the sources), J (the destinations) and K (the price zones) is at least 20. In these cases,

one can also see a remarkable difference in the calculation times. With the increasing cardinality of the sets the differences increase.

Fig. 3 depicts the influence of the calculation times on the cardinality of I, J and K (note that the cardinality of all the sets is the same and is denoted S). We can see that the calculation times seem to increase exponentially. The orange curve corresponds to the model with $x_{ijk} \in R_0^+$, the blue curve to the model with $x_{ijk} \in Z_0^+$.

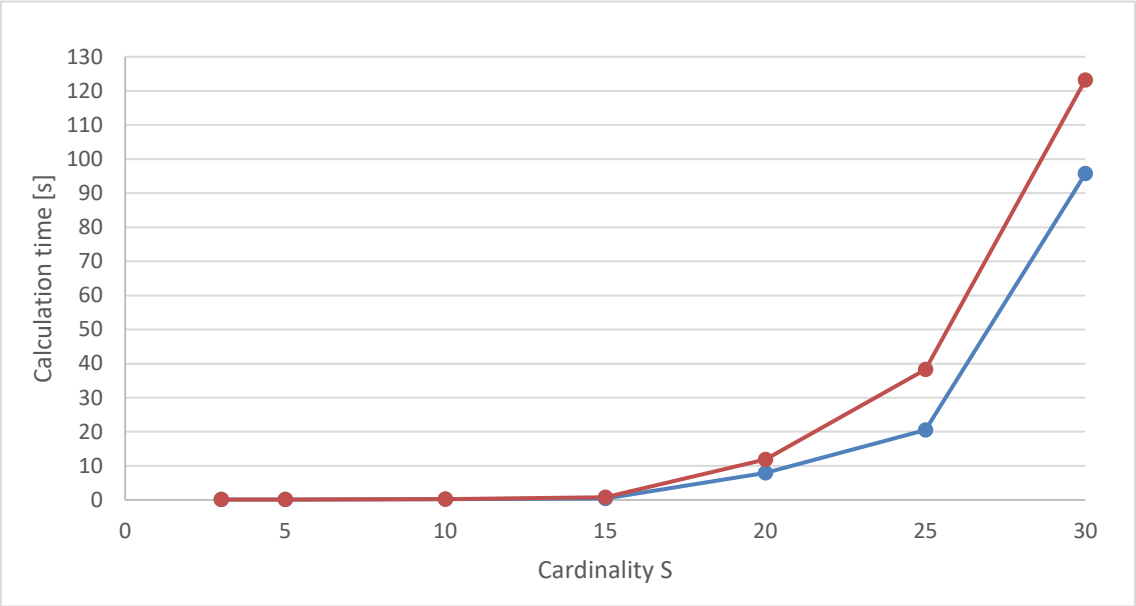


Figure 3 Impact of the cardinality S on the calculation times

Fig. 4 shows how the differences in the calculation times for the models with the different domain of definition of x_{ijk} depend on the cardinality S.

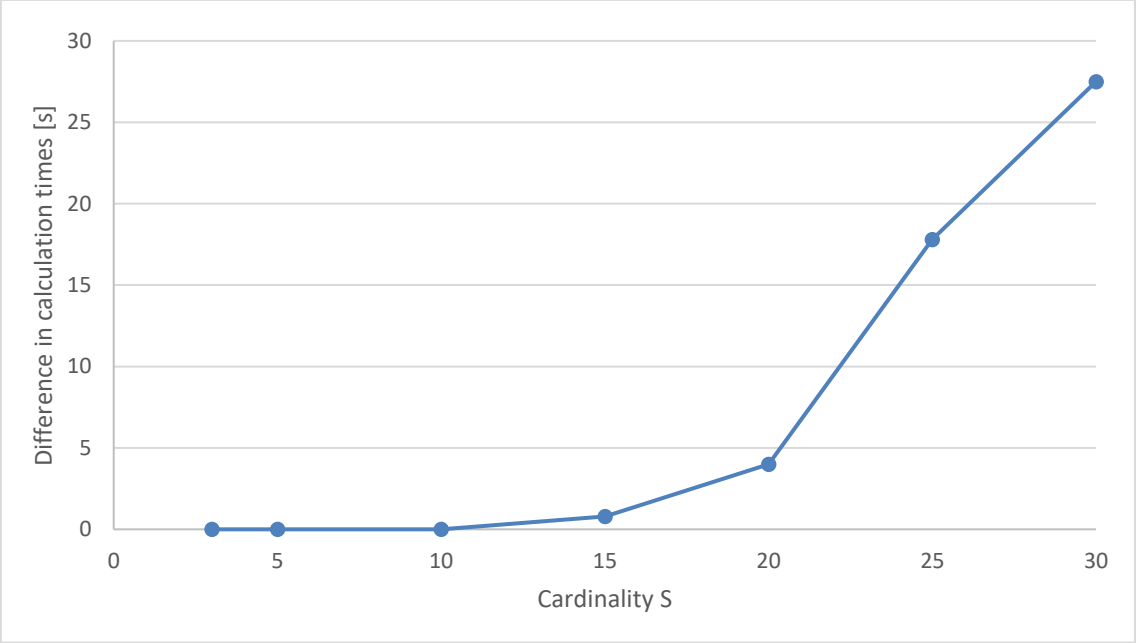


Figure 4 The dependence of the calculation times on cardinality S

5 Conclusions

The article is focused on the modification of the transportation problem model. The modification is distinctive in the fact that the objective function is nonlinear, more precisely is step-wisely decreasing. The mathematical model which is presented in the article consists of the optimisation criterion and 8 groups of the constraints. The variables

x_{ijk} model how many units of goods is transported from the source $i \in I$ to the destination $j \in J$ within the price zone $k \in K$. The mathematical model enables using two possible domains of definition of x_{ijk}. The variables can take the nonnegative values R_0^+ or the nonnegative integers Z_0^+ only. The second group of the variables denoted u_{ijk} decides about assigning the price zone $k \in K$ to transportation between $i \in I$ and $j \in J$. The mathematical model consists of $m \cdot n \cdot (4 \cdot p + 1) + m + n$ constraints and $2 \cdot m \cdot n \cdot p$ variables ($m \cdot n \cdot p$ variables are binary).

The presented mathematical model was tested using some examples. In total, 14 experiments in optimisation software Xpress-IVE were carried out. The experiments were focused on finding out how the calculation time is influenced with the number of the sources, destinations and price zones. The experiments confirmed the fact that the calculation time increases exponentially for both possible domains of definition of x_{ijk}. Despite the great number of the variables the calculation times were relatively short. We can suppose that it is because of applied simplification of the modelled problem – for each direction from $i \in I$ to $j \in J$ the width of the price zones and the coefficients of quantity discount were assumed to be equal. Another piece of knowledge we found out was that the calculation times are greater if x_{ijk} can take the nonnegative values R_0^+.

About our future research, we would like to test the model with a more complex system of the price zones; within this research we would like to aim at the differences in the calculation times for both possible domains of definition. Another possible stream of our research is to consider several carriers offering transportation; the individual carriers can differ in the transportation costs and the offered quantity discounts.

References

[1] Hitchcock, L.: Distribution of a Product from Several Sources to Numerous Localities. *Journal of Mathematical Physics*, vol. 20, 1941.

[2] Dantzing, G., B.: Application of the Simplex Method to a Transportation Problem.

[3] Shen, S., Dechen, Z., Xiaofei, X.: Genetic Algorithm for the Transportation Problem with Discontinuous Piecewise Linear Cost Function. *International Journal of Computer Science and Network Security*, vol. 6, No. 7A , 2006.

[4] Janáček, J.: Matematické programování. EDIS, 1999. ISBN 80-7100-573-8, 225 p.

[5] Janáček, J.: Operační analýzy II. ALFA, 1983. ISBN 63-724-83, 142 p.

[6] Goľštejn, J., G.; Judin, D., B.: Zadači linejnogo programmirovanija transportnogo tipa. Moskva: NAUKA, 1969

[7] Cenek, P., Jánošíková, Ľ. Models and Optimisation in Transports and Logistics. Žilina: EDIS, 2008. ISBN 978-80-8070-951-8

Effectiveness as a New Focus of the MAVT MCDM Methods

Roman Guliak[1]

Abstract. Having addressed the methods based on the multi-attribute utility theory (MAUT), we defined that the overall assessment of the alternative is focused only on its quantitative aspect taking into account absolute values of the criteria and giving the value assessment to each alternative. In other words, the level of development (performance) or the efficiency ("doing things right") of an alternative are measured, when the effectiveness is left beyond the attention. In spite of frequent usage of this term in the literature and acknowledged importance in management practice, the well-grounded methodology for the effectiveness has been missing. The generic view of effectiveness explained as "doing right things" underpins the suggested methodology. Basing on this, the focus in the measurement is shifted from the weighted absolute values of criteria to the weighted inter-criteria proportions. The new Difference Additive Weighting method is suggested to solve the problem of utility dependence during proportions investigation. This method is looking for the alternative with the "effective design" or "effective performance" tracing the degree of relative targets attainment with respect to the priority vector and level of uniformity

Keywords: effectiveness, goal attainment, multi-attribute, MCDM, SAW, difference approach, utility independence.

JEL Classification: C44, R11, R58
AMS Classification: 91B16, 91B82

1 Introduction

The "effectiveness" is the vague concept originally appeared in the managerial context of organizational performance measurement. There were many debates devoted to this concept, which successfully have been finished with some consensus bounding the set of accepted approaches being appropriate for specific situation [2]. Despite high interest and variety of the conceptual approaches [8, 10, 11, 12] to effectiveness, no one to the best of our knowledge was quantified and embedded into by the certain methodology. This fact is the main motivation of this paper.

On the contrary, all efforts are focused on the measurement of other aspects of performance. For instance the term "efficiency" is used sometimes as a synonym to "effectiveness". In this case the criteria follow should the input-output structure and nonparametric Data Envelopment Analysis (DEA) is applied for the efficiency measuring [3]. Both these terms reflect two different complementary aspects revealing the performance of the alternative. If the "efficiency" is about "doing things right", the "effectiveness" is about "doing right things" [1]. Meanwhile another generic and less specific approach exists. Its focus is related to the bald "having things done" without any other additional aspects of analysis. Normally this approach is applied to measure different synthetic properties, such as "level of development", "quality of life", etc. It does not require any relations between its substitutes and Simple Additive Model (SAM) [6] from multi-attribute value theory (MAUT) theory can be appropriately used.

It is necessary to mention, that the performance context is not the only one appropriate for the effectiveness exploitation. In the broadest sense, the effectiveness can be analyzed from passive and active points of view. Active is about how an alternative performed, passive – how an alternative should be designed. Having stressed this point, the wide application of this term discovering the aspect "doing right things" in different fields is unquestionable. Moreover, it is fair to say, that any analysis of the performance following both directions, for example efficiency via effectiveness, always tends to be more comprehensive and informative.

Considering the fact of the variety of conceptual approaches to the effectiveness measurement, the most discussed and widely used remains the "Goal attainment" approach [7] focused on the ends exclusively, namely on the achievement of goals, objectives, targets, etc. [8]. According to this approach, the effectiveness is defined as degree to which targets achieved [4]. In our research, this definition is the starting point for the development of

[1] VŠB - TU Ostrava, Faculty of Economics, Sokolská třída 33, 70121 Ostrava 1, e-mail: roman.guliak@vsb.cz; O. M. Beketov National University of Urban Economy in Kharkiv, 17, Marshal Bazhanov Street, Kharkiv, 1002, Ukraine.

the method for the effectiveness measurement. Based on this the aim of the paper is development of the method for the effectiveness measurement based on "Goal attainment" approach.

2 Multi-Attribute Utility Theory Platform for effectiveness measurement

The effectiveness as a multi-dimensional phenomenon requires some aggregating function to be processed. The Multi-Attribute Utility Theory (MAUT) [9] is the most suitable platform to start with. The most widely used method in MCDM is so far the Simple Additive Weighting method (SAW) [6]. It provides the most intuitive framework easy dealing with multiple criteria problems:

$$U_k = \sum_{j=1}^{n} w_j r_{kj}(x), \tag{1}$$

where U_k is the utility of the *k*-th alternative and w.r.t. *j*-th criterion; w_j – weight of *j*-th (*j* = 1,..., *n*) criteria under the condition of $\sum_{j=1}^{n} w_j = 1$; $r_{kj}(x)$ – normalized value of actual value *x*, in terms of the *k*-th alternative ($k = 1,...,m$) and *j*-th criterion (*j* = 1,..., *n*).

Verbally, a utility value can be perceived as a representation of decision maker's preferences over alternatives, with one preferred to another if and only if its expected utility is greater. The best alternative a^* is defined as:

$$a^* = \left\{ U_k(x) \middle| \max_k U_k(x) \right\}. \tag{2}$$

The linear additive function correctly integrates the preferences of decision makers into a total utility value only when the independence assumptions [9] are being met. If we focus just on the absolute level of targets and their importance, we easily can apply the SAW method and measure, for example, aggregated indicators such as "level of development", "synthetic performance", "quality of life". In real world problems, mutual independence preference condition is held quite rarely and simply very often ignored by decision makers.

3 Pre-conditions for the method development

The springboard of the further analysis is the basic definition of the effectiveness. As stated above, according to the "goal attainment" approach, the effectiveness is the degree to which targets achieved. We will slightly enrich this ultimately simple definition, making it more precise. The effectiveness is the relative degree to which targets (uniformly) achieved with respect to the priority vector. Relying on this working definition, the following pre-conditions are listed and should be considered in the method:

1. The focus is on the quality of performance (or design). Effectiveness is a relative phenomenon, expressing the concept of "doing right things" with the emphasis on the character of the "things". It is not just benchmarking of a performance, aggregating attributes of alternatives. It is the comparison between "things" (targets) being more or less right according to the DM's preferences.
2. Comparison between target attainments, not between absolute values of initial criteria. The Effectiveness measurement method is aimed at the analysis of proportions.
3. No compensation between absolute values. Targets that are more preferential have a higher priority in an evaluation. Higher value of the less important target attainment cannot compensate the lower value of the more important one as it is allowed in the Simple Additive Sum method.
4. The degree of the right target attainment is measured. The targets of alternative are rightly fulfilled, when the proportions between target attainments are defined in according to the vector of weights.
5. Uniformity could be important (optionally). Fragmentary following the right proportions does not allow in a full extent to explore the effectiveness phenomenon in a full extent. The uniformity of targets achievements could be a constituent of the effectiveness concept and another aspect to be considered.

4 Difference decomposition for the utility independent proportions' measurement

The MCDM problem is described as follows: initial set of alternatives $A = \{a_k | k = \overline{1,m}\}$, set of criteria $C = \{c_j | j = \overline{1,n}\}$ characterizing the performance (design) of alternatives, performance matrix $X = [x_{kj}]_{m \times n}$, showing all values assigned to the alternatives relating to an each criterion and weights of criteria

denoted by $W = \{w_j | j = \overline{1,n}\}$. Taking SAW method as a basis for the effectiveness measurement, the utility independence is going to be violated. It happens because the calculation of shares (or proportions) of target achievement levels creates dependency between single utility functions. Thus, the total utility E_k is not equal to the sum of single utility functions e_{kj} :

$$E_k(x_{k1},...,x_{kn}) = f[e_{k1}(x_{k1}),...,e_{kn}(x_{kn})] \neq \sum_{j=1}^{n} w_j \cdot e_{kj}(x_{kj}) \quad (3)$$

To solve this problem and avoid using of non-additive value function, the transformation of the SAW has to be done. Let us suppose that original criteria set $C = \{c_j | j = \overline{1,n}\}$ is decomposed to get new extended superset $\bigcup_{j=2}^{n} C_{\tau(j)>\tau(v)}$, where $\tau(j), \tau(v)$ is the permutation on W, that is $w_{\tau(1)} < w_{\tau(j)} < w_{\tau(n)}$ for $j, v = \overline{1,n}$. Decomposition implies *n* criteria to be extended into the *n*-1 subsets of sub-criteria composing the superset $\bigcup_{j=2}^{n} C^*_{\tau(j)>\tau(v)} = C_{\tau(2),\tau(v)} \cup ... \cup C_{\tau(j),\tau(v)} ... \cup C_{\tau(n),\tau(v)} | j, v = \overline{1,n} : j > v$ derived from the lower triangular matrix $(C \times C)$ (table 1). Its elements $c_{\tau(j)/\tau(v)}$ (the same as $c_{\tau(j)>\tau(v)}$) denote preference relations between *j*-th (more important) and *v*-th (less important) criteria.

$c_{\tau(j)}$ \ $c_{\tau(v)}$	$> c_{\tau(1)}$	$> c_{\tau(2)}$	$> c_{\tau(3)}$	$> c_{\tau(4)}$	$> c_{\tau(5)}$
$c_{\tau(1)} >$	–	–	–	–	–
$c_{\tau(2)} >$	$c_{\tau(2)/\tau(1)}$	–	–	–	–
$c_{\tau(3)} >$	$c_{\tau(3)/\tau(1)}$	$c_{\tau(3)/\tau(2)}$	–	–	–
$c_{\tau(4)} >$	$c_{\tau(4)/\tau(1)}$	$c_{\tau(4)/\tau(2)}$	$c_{\tau(4)/\tau(3)}$	–	–
$c_{\tau(5)} >$	$c_{\tau(5)/\tau(1)}$	$c_{\tau(5)/\tau(2)}$	$c_{\tau(5)/\tau(3)}$	$c_{\tau(5)/\tau(4)}$	–

Table 1 Extension of an initial criterion set.

Altogether subsets count (n^2-n)/2 new benefit paired sub-criteria. Table 1 gives an example of new extended superset $\bigcup_{j=2}^{5} C^*_{\tau(j)>\tau(v)}$ derived from initial $C = \{c_j | j, v = \overline{5,n}\}$ by difference decomposition, when $j > v$. Having defined the weights of all initial criteria, they are placed in the rows and columns such as $w_{\tau(1)} < w_{\tau(j)} < w_{\tau(n)}$ for $j, v = \overline{1,n}$. At the intersection of each row and column we obtain differences between more important criteria (minuends) and less important ones (deductions). All these intersections give the platform for the new paired sub-criteria and new sub-weights. All the new benefit paired sub-criteria are presented by $n-1$ subsets $\bigcup_{v=1}^{j-1} C_{\tau(j)>\tau(v)} = \{c_{\tau(j)/\tau(1)},...,c_{\tau(j)/\tau(v)},...,c_{\tau(j)/\tau(j-1)}\}$ with $j = \overline{2,n} : j > v$. For example, from the table 1 we see that $c_{\tau(2)}$ is more important than $c_{\tau(1)}$, consequently from the intersection we have new paired sub-criterion $c_{\tau(2)/\tau(1)}$. Having decomposed the set of original criteria, the normalised performance matrix $L = [l_{kj}]_{m \times n}$ (see form. 7) and weights $W = \{w_j | j = \overline{1,n}\}$ are respectively decomposed for the extension in a similar way as in table 1, but using difference operator:

$$l_{\tau(j)/\tau(v)} = l_{\tau(j)} - l_{\tau(v)}, \text{ the same is for } w_{\tau(j),\tau(v)} = w_{\tau(j)} - w_{\tau(v)}. \quad (4)$$

For example, from the section 6 we know that criteria weights have been defined such as $w_2 > w_1$, meaning higher importance of gross fixed capital formation (c_2 – minuend) over economically active population (c_1 – deduction). It means that the higher priority in targets' achievement should be given to the c_2. The weight of the new obtained sub-criterion $c_{2,1}$ is $w_{2,1}$= 0.07–0.04= 0.03. If $l_{\tau(2)/\tau(1)}$ for the 1st region is lower than for the 2nd one, the latter has better (more effective) proportions, even though absolute values of c_1 and c_2 from the 1st region are

higher. In other words, 2nd region achieves targets better or does "right things" in terms of etalon proportions established by weights. To remind one more time, just proportions are the subject for the measurement in the frame of the effectiveness analysis.

The benefit subset is the sum of differences between minuend and deductions. However, the multiple participation of the minuend is accompanied by its cost participation as well, when it plays the role of deduction in the subset initiated by other more important criteria, unless it is the last *n*-th the most important criterion. By such decomposition the participation of each original criterion is accounted from both benefit and cost sides separately. Thus, the witnessed in (1) utility dependence and dual nature (cost and benefit) of original criteria can be avoided due to the decomposition and formation of new paired sub-criteria.

5 Difference Additive Weighting method

Based on the preconditions from the subsection 3 the following targets should be reached by the developed method: 1. to catch the relative proportions of target attainments; 2. to measure the degree of right target attainment; 3. to consider the uniformity target attainment degree as an additional factor of effectiveness (optionally). The methodological contribution of the proposed Difference Additive Weighting method (DAW) for the effectiveness measurement is steamed out from the difference decomposition approach introduced in the section 4 and embedded into the SAW method frame [6] presented in the section 2. Further we present steps of the suggested new DAW method:

1. Determine the initial set of alternatives $A = \left\{ a_k \middle| k = \overline{1,m} \right\}$.
2. Select the original criteria $C = \left\{ c_j \middle| j = \overline{1,n} \right\}$ and decompose it to the extended set of sub-criteria $\bigcup_{j=2}^{n} C^*_{\tau(j)>\tau(v)}$ for $j,v = \overline{1,n} : j > v$, where $\tau(j),\tau(v)$ is the permutation on W, that is $w_{\tau(1)} < w_{\tau(j)} < w_{\tau(n)}$.
3. Form the performance matrix $X = [x_{kj}]_{m\times n}$ showing all values assigned to the alternatives relating to an each criterion.
4. Obtain weights of criteria denoted by $W = \left\{ w_j \middle| j = \overline{1,n} \right\}$ and decompose it using difference function to obtain new extended *n*-1 subsets $\overset{j-1}{\underset{v=1}{W}}_{\tau(j)/\tau(v)}$ for $j = \overline{2,n} : j > v$ with sub-weights $w_{\tau(j)/\tau(v)}$ calculated by

$$w_{\tau(j)/\tau(v)} = w_{\tau(j)} - w_{\tau(v)}. \tag{5}$$

5. Normalise the new extended sub-weights $w_{\tau(j)/\tau(v)}$ to obtain the sum of $w^*_{\tau(j)/\tau(v)}$ equal to 1

$$w^*_{\tau(j)/\tau(v)} = \frac{w_{\tau(j)} - w_{\tau(v)}}{\sum_{j=2}^{n}\sum_{v=1}^{j-1}(w_{\tau(j)} - w_{\tau(v)})}. \tag{6}$$

6. Normalise initial performance values x_{kj} by calculating the level of target achievement l_{kj} for the *k*-th alternative and *j*-th criterion

$$l_{kj} = \frac{x_{kj} - x_j^{\min}}{x_j^{\min}} \text{ and for the cost criteria } l_{kj} = x_j^{\max} / x_{kj} - 1 \tag{7}$$

where $x_j^{\min}$ ($x_j^{\max}$) – minimum (maximum) value for *j*-th criterion among *m* alternatives; $x_j^{\min} \geq 0$ as we make an assumption that targets are always achieved at some minimal level.

7. Decompose levels of target achievement l_{kj} using difference function to obtain new extended target achievements $l_{k,\tau(j)/\tau(v)}$, encompassing relations between two criteria:

$$l_{k,\tau(j)/\tau(v)} = l_{k,\tau(j)} - l_{k,\tau(v)} \text{ for } j = \overline{2,n} : j > v. \tag{8}$$

8. Standardise decomposed levels of target achievement to obtain their z-values $l^*_{k,\tau(j)/\tau(v)}$ for $\sigma(j)/\sigma(v)$-th criterion among *m* alternatives:

$$l^*_{k,\tau(j)/\tau(v)} = \frac{l_{k,\tau(j)/\tau(v)} - \bar{l}_{\tau(j)/\tau(v)}}{S_{\tau(j)/\tau(v)}}, \tag{9}$$

where $\bar{l}_{\tau(j)/\tau(v)}$ – average value of target achievement, $S_{\tau(j)/\tau(v)}$ – standard deviation of target achievement.

9. Calculate the effectiveness E_k representing the decision maker's preferences over *k*-th alternative considering the extended weighted target achievements

$$E_k = \sum_{j=2}^{n} \sum_{v=1}^{j-1} l^*_{k,\tau(j)/\tau(v)} \times w^*_{\tau(j)/\tau(v)} \text{ for } j = \overline{2,n} : j > v . \tag{10}$$

10. Calculate the coefficient of variation CV_k of extended weighted target achievements $lw_{k,\tau(j)/\tau(v)}$ to measure their uniformity

$$CV_k = \frac{\sigma(lw_{k,\tau(j)/\tau(v)})}{\overline{lw}_{k,\tau(j)/\tau(v)}} \text{ for } k = 1,\ldots,m, \tag{11}$$

$$\text{when } lw_{k,\tau(j)/\tau(v)} = l^*_{k,\tau(j)/\tau(v)} \times w^*_{\tau(j)/\tau(v)}, \tag{12}$$

where $\overline{lw}_{k,\tau(j)/\tau(v)}$ – average value of weighted target achievements, $\sigma(lw_{k,\tau(j)/\tau(v)})$ – standard deviation of weighted target achievements.

11. Calculate the balanced E_k^b by means of CV_k as follows:

$$E_k^b = \frac{E_k}{CV_k}. \tag{13}$$

6 Effectiveness of regions from Visegrad group

The suggested DAW method was applied for the data describing the effectiveness of performance of 35 Visegrad NUTS 2 regions in 2013 year. This data was chosen for the effectiveness measurement because it is the most complex and latest data available from the Eurostat database [5]. The effectiveness measurement is combined with regional productivity (fig. 1) to show that there is no correlation between them and both concepts are considered to be nothing but complements in analysis. The set of criteria consists of the following original 7 indicators and their weights reflecting the preferences in the following development pattern: gross domestic product (PPS) (w_7 = 0.25); unemployment (persons) (w_6 = 0.21); employment (persons) (w_5 = 0.18); human resources in science and technology (persons with tertiary education and/or employed in science and technology) (w_4 = 0.14); total intramural R&D expenditure (euro) (w_3 = 0.11); gross fixed capital formation (euro) (w_2 = 0.07); economically active population (persons) (w_1 = 0.04). Weights of criteria were established by Ratio Method.

Due to the decomposition 7 original indicators have been extended to the 21 sub-criteria. The set of weights as well as set of criteria was extended by (4). Having applied formulae 5-13 balanced Effectiveness was measured. Uniformity aspect is taken into account on regional level as all targets has to be achieved evenly and the high level of compensation is not acceptable, that is why this case would be negatively influenced by the coefficient of variation (form. 13).

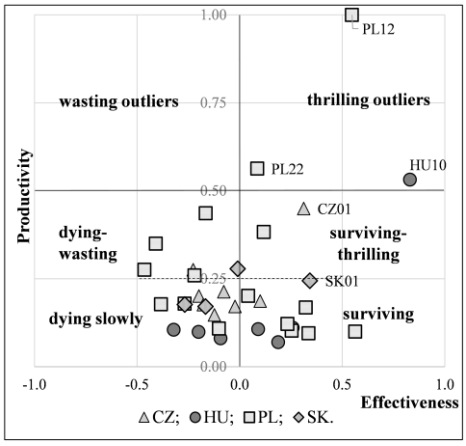


Figure 1 Productivity-Effectiveness matrix

For the analysis of regional performance we have chosen two-dimensional space, which includes Productivity on axis Y (measured as ratio of Gross Value Added / working hours) and Effectiveness on axis X. This space presented as Productivity-Effectiveness matrix describes regional performance in four main quadrants. In our case

we mixed some quadrants because of the presence of outliers in the data. They have got high productivity and high effectiveness, what makes them leaders within Visegrad group with balanced performance. As well we highlighted two mixed areas showing regions marked as "dying-wasting" and "surviving-thrilling". Regions under the "dying-wasting" mark are considered as highly productive but not effective. In comparison to "surviving-thrilling" their resources are not used in a full extent and targets need to be re-established according to the development pattern reflected in the weights of initial criteria. Thus "surviving-thrilling" regions reach the "right targets" and they are quite successful in it what is proved by the relatively high productivity. Special attention should be given to regions falling to the opposite "dying slowly" sector as they do not possess enough resources and moreover spend them in not effective way. Meanwhile, better place in the matrix take more promising "surviving" regions due to their higher level of effectiveness stressing the right path of development but still low level of resources reflected in low productivity.

Conclusion

Even though effectiveness originates from a plenty of conceptual approaches, none of them was quantified by some grounded methodology. The investigation of criteria proportions for the measurement of level of targets' attainment causes the violation of utility independence and as a result the SAW method fails. To avoid this problem the DAW method is suggested. Its main idea is the decomposition of an original set to obtain an extended set of sub-criteria released from dependency. All newly obtained sub-criteria are treated as independent benefit ones contributing to the total utility function.

One should mention that it is unlikely the case when effectiveness is the only focus of the analysis. The effectiveness investigation should just complement the classical analysis of performance level or efficiency measurement. Concept of effectiveness plays the role of a source giving the additional information and broadening modelled picture of the reality as it was pictured on the example of development of Visegrad regions.

Further analysis needs to account the interaction between criteria. This will allow their proper treatment considering positive and negative synergies resulting from complements and substituents respectively.

Acknowledgements

This article has been funded by INFINITY project in the framework of the EU Erasmus Mundus Action 2 and by Student Grant Competition project SGS SP 2017/125 of Faculty of Economics, VŠB-Technical University of Ostrava.

References

[1] Anthony, R. N., Govindarajan, etc.: *Management Control Systems*. McGraw-Hill Education - Europe, London, United States, 2014, 664 p.
[2] Cameron, K.S.: The effectiveness of ineffectiveness, in Staw B.M and Cummings, L.L. *Research in Organisational Behaviour*, 1984, 6, JAI Press, Greenwich, CT, 235-85.
[3] Charnes A., Cooper, W. W. Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research*, Vol. 2 (1978), No. 4, pp. 429–444.
[4] Etzioni, A.: *Modern organization*. Englewood Cliff, New Jersey, Prentice-Hall, 1964, 120 p.
[5] Eurostat database: available from: http://ec.europa.eu/eurostat.
[6] Fishburn, P.C.: Additive Utilities with Incomplete Product Set: *Applications to Priorities and Assignments, Operations Research Society of America (ORSA) Publication*, Baltimore, MD, 1967.
[7] Goodman, P. S., Pennings, J. M. (eds.): *New perspectives on organizational effectiveness*. San Francisco - London, Jossey-Bass Publishers, 1977.
[8] Henri, J.F.: "Performance measurement and organizational effectiveness: bridging the gap". *Managerial Finance*, Vol. 30 Issue: 6 (2004), pp.93-123.
[9] Keeney, R.L., Raiffa, H.: *Decisions with multiple objectives–preferences and value trade-offs*. Cambridge University Press, Cambridge & New York, 1993, 569 p.
[10] Love, P. and Skitmore, M.: Approaches to Organisational Effectiveness and Their Application to Construction Organisations. In Thorpe, A., Eds. *Proceedings 12th Annual Conference and Annual General Meeting*, 1996.
[11] Mette A, Joseph, C. P., David, N. Reese and Fai, T.: Measuring overall efficiency and effectiveness using DEA. *European Journal of Operational Research*, vol. 178, issue 1 (2007), p. 305-321.
[12] Oghojafor, B.E.A., Muo, F.I., Aduloju, S.A.: Organisational Effectiveness: Whom and What Do We Believe? *Advances in Management & Applied Economics*, vol.2, no. 4 (2012), p. 81-108.
[13] Price, J.L.: Organisational effectiveness. Homewood, Ill, 1968: Richard D. Irwin, Price, J.L. (1972). *The Study of Organisational Effectiveness, Sociological Quarterly*, 13(3) 3-15.

Contribution to Economic Efficiency Evaluation of Projects in Terms of Uncertainty

Simona Hašková[1]

Abstract. The research subject is the profitability of investments into production systems with robust linkages between inputs and outputs, working in conditions of uncertainty on the input side. Uncertainty is understood as a state of vagueness that does not provide a decision-maker with any guidance to construct rational estimates; only possible ranges of input values are known. A typical representative of such a system is a solar photovoltaic power plant, by means of which the research findings are demonstrated. The basic idea of the herein presented approach is the thesis that in terms of uncertainty there is a much greater chance to correctly predict the intervals at which the respective point values are found than to correctly estimate a single point value. Within this argument the relevant formulas are translated from the language of arithmetic to the language of intervals. In it the significant points of the interval of profitability values are derived, which, in addition to standard information about the profitability of investments (standard IRR), also provide information about the measure of investment safety (in terms of resistance against possible financial loss) and the measure of hope for a premium assessment. These conclusions are the major original contribution.

Keywords: investment profitability, uncertainty, internal rate of return, interval estimate, solar photovoltaic power plant.

JEL Classification: C51
AMS Classification: 90B90

1 Introduction

There are a number of criteria for the evaluation of a business plan (project) in economic theory and practice. Some of them are based only on evaluation of economic efficiency (monocriterial decision) measured by various methods (see e.g. Brealey et al. [1]). In the case of long-term projects the most important criteria of monocriterial decision-making are the criteria based on the calculation of the net present value (NPV) and the calculation of the internal rate of return (IRR), if this exists, and thus can be determined. Taking into account more criteria simultaneously, we talk about multi-criteria decision making (see [3]). From the perspective of game theory this is considered in both cases as "a game against nature". If uncertainty is connected with the result, it usually has two main sources: the vagueness of the rules according to which it is played, and external circumstances that cannot be influenced by the decision-maker (a combination of both cannot be ruled out). In multi-criteria decision-making vagueness usually prevails, which is associated with the need to project the results of the evaluation according to various criteria onto a common scale to which they can be added (see [10]). One of the options of how to handle this vagueness is shown in [11].

The aim of this paper is to demonstrate how it can be tackled, at least partially, if circumstances cannot be influenced. If it is possible to describe the occurrence of the consequences of the source of uncertainty by means of a statistical method (e.g. by the known probability distribution), we take such uncertainty as risk [5]. If it is impossible to describe it, we call the uncertainty vagueness. Thus, for example, NPV determined by calculation according to equation (1), see section 2, will almost certainly differ from the actual increase in investor wealth generated by an investment in the project. This is partly because of the fact that the relationship (1) is based on mathematical models of rather vague ideas that the more distant or risky a payment is, the less significant from the current perspective. However, the factors of time and risk are not nearly as robust as physical laws (this was shown in [2] on evolution of the return rates in dependence on the reduction rates of the photo-solar systems purchasing cost, which is difficult to predict). The greater part of the difference between the result of calculation and reality will be at the expense of uncertainty associated with the unfamiliarity of future exact points of the cash flow (CF_i) values and the discount rate (r_j). There is a much greater chance to correctly predict the intervals at which the respective point values of these variables will be located than to correctly estimate these single point values. We will be more successful if we derive from ranges of possible values instead of uncertain point values in calculating the NPV.

[1]The Institute of Technology and Business/The School of Expertness and Valuation, Okružní 517/10, České Budějovice, Czech Republic, e-mail: haskovas@post.cz.

This is the basic idea of the approach presented here, the inspiration of which became the works [6] and [7], which is developed in the next section in detail. First, the equation (1) effective for point values NPV, CF_i and r_j is reformulated to the relationship (2) effective for intervals **NPV**, $\mathbf{CF_i}$ and $\mathbf{r_j}$ of their possible (resulting, observed or entered) values. Within the calculus of intervals that are represented by their significant points (3), whose arithmetic operations **(+)**, **(–)**, **(·)** and **(/)** are a natural extension of the operations +, –, · and / of the arithmetic of real numbers, the significant points of **NPV** interval are derived from the important points of intervals $\mathbf{CF_i}$ and $\mathbf{r_j}$. They tell us far more about the effectiveness of the project than the calculated point NPV value from the equation (1).

Introducing the simplified, but quite realistic assumptions allowing the existence of IRR, we then derive the conditions for calculating the significant points of **IRR** from the relationship describing the significant points of **NPV**. They are applied in section 3 dedicated to practical application, the task of which is to compare the investment profitability of two solar photovoltaic power plants (FVE) of the same type put into operation at different times in the Czech Republic.

2 Methodological approach

Net Present Value (NPV) of the project for the lifetime *n* years is generally defined by the relation

$$\mathrm{NPV} = \sum_{i=0}^{n} \mathrm{CF_i} \,/ \prod_{j=1}^{i} (1 + \mathrm{r_j}), \tag{1}$$

where
CF_0 is the initial capital expenditure (negative payment),
CF_i, $i > 0$ are positive or negative payments generated by the project in the i-th year of its lifetime,
r_j is a positive discount rate per annum valid in the j-th year of the project lifetime.

In the first step of our approach to the NPV calculation in the case of unknown uncertain values CF_i and r_j we substitute into (1) for NPV, CF_i and r_j the symbols of triads of real numbers $\mathbf{NPV} = (NPV_l, NPV, NPV_r)$, $\mathbf{CF_i} = (CF_{il}, CF_i, CF_{ir})$ and $\mathbf{r_j} = (r_{jl}, r_j, r_{jr})$, composed of the significant points of files (intervals) of possible calculated, given or observed uncertain values. The left (index l), respectively right (index r) marginal number in the respective trio provides the smallest, respectively the greatest element of the set. The middle number indicates the value of the most common and most anticipated element (it usually represents the number that would be estimated within the standard approach in terms of risk).

Likewise, we replace the operations +, –, · and / of arithmetic of real numbers with operations **(+)**, **(–)**, **(·)** and **(/)** applicable to intervals. As a result of the said replacements the relationship (1) transforms to the form

$$\begin{gathered}\mathbf{NPV} = \mathbf{CF_0}(+)[\mathbf{CF_1}(/)(1(+)\,\mathbf{r_1})](+)\ldots(+)\left[\mathbf{CF_i}\,(/)\,\left((1(+)\mathbf{r_1})(\cdot)\ldots(\cdot)(1(+)\mathbf{r_i})\right)\right] \\ (+)\ldots(+)\left[\mathbf{CF_n}(/)\left((1(+)\mathbf{r_1})(\cdot)\ldots(\cdot)(1(+)\mathbf{r_n})\right)\right].\end{gathered} \tag{2}$$

Let $\mathbf{A} = (A_1, A, A_r), \mathbf{B} = (B_1 B, B_r)$, k is a constant.
Then for operations of calculus of intervals this applies:

- k **(+) A** = $(k + A_l, k + A, k + A_r)$,
- **(–) A** = $(-A_r, -A, -A_l)$,
- **A (+) B** = $(A_l + B_l, A + B, A_r + B_r)$,
- **A (–) B** = $(A_l - B_r, A - B, A_r - B_l)$,
- If all relevant points of intervals including the constant k are positive, then
- k **(·) A** = $(k \cdot A_l, k \cdot A, k \cdot A_r)$,
  A (·) B = $(A_l \cdot B_l, A \cdot B, A_r \cdot B_r)$ a
  A (/) B = $(A_l / B_r, A / B, A_r / B_l)$.

(3)

By translation of equality (2) from the language of the calculus of intervals into the language of calculus of significant points according to the rules of the system (3) we obtain, after adjustments:

$$\begin{gathered}\mathrm{NPV_1} = \mathrm{CF_{01}} + \sum_{i=1}^{n} [\max(\mathrm{CF_{i1}}, 0) / \prod_{j=1}^{i} (1 + \mathrm{r_{jr}}) + \min((\mathrm{CF_{i1}}, 0) / \prod_{j=1}^{i} (1 + \mathrm{r_{j1}})], \\ \mathrm{NPV} = \mathrm{CF_0} + \sum_{i=1}^{n} \mathrm{CF_i} \,/ \prod_{j=1}^{i} (1 + \mathrm{r_j}),\end{gathered} \tag{4}$$

$$\mathrm{NPV_r} = \mathrm{CF_{0r}} + \sum_{i=1}^{n} [\max(\mathrm{CF_{ir}}, 0) / \prod_{j=1}^{i} (1 + \mathrm{r_{j1}}) + \min((\mathrm{CF_{ir}}, 0) / \prod_{j=1}^{i} (1 + \mathrm{r_{jr}})].$$

Assuming $\mathbf{r_j} = \mathbf{r}$ for all j we obtain $(r_{jl}, r_j, r_{jr}) = (r_l, r, r_r)$, and therefore $\prod_{j=1}^{i}(1 + r_{j1}) = (1 + r_1)^i$, $\prod_{j=1}^{i}(1 + r_j) = (1 + r)^i$ and $\prod_{j=1}^{i}(1 + r_{jr}) = (1 + r_r)^i$. If moreover $CF_{il} \geq 0$ for all $i > 0$, then, after taking into account all these facts in relations of system (4) this becomes

$$\begin{aligned} \mathrm{NPV_1} &= \mathrm{CF_{01}} + \sum_{i=1}^{n} \mathrm{CF_{i1}} / (1 + \mathrm{r_r})^i, \\ \mathrm{NPV} &= \mathrm{CF_0} + \sum_{i=1}^{n} \mathrm{CF_i} / (1 + \mathrm{r})^i, \\ \mathrm{NPV_r} &= \mathrm{CF_{0r}} + \sum_{i=1}^{n} \mathrm{CF_{ir}} / (1 + \mathrm{r_1})^i. \end{aligned} \tag{5}$$

The system (5) allows formulating conditions for the calculation of significant points IRR_l, IRR and IRR_r of **IRR** interval that can be comfortably solved by the computer tools as shown in [4] in the form:

$$\begin{aligned} \sum_{i=1}^{n} (\mathrm{CF_{i1}} / (1 + \mathrm{IRR_1})^i) &= -\mathrm{CF_{01}}, \\ \sum_{i=1}^{n} (\mathrm{CF_i} / (1 + \mathrm{IRR})^i) &= -\mathrm{CF_0}, \\ \sum_{i=1}^{n} (\mathrm{CF_{ir}} / (1 + \mathrm{IRR_r})^i) &= -\mathrm{CF_{0r}}. \end{aligned} \tag{6}$$

3 Application part: profitability analyses of FVE investments in terms of uncertainty

The demand for solar photovoltaic power plants (FVE) of various types increased sharply in the Czech Republic (CR) due to generous subsidies for environmentally clean sources of electricity during the years 2008 - 2010. This increase stimulated further technological development, which, together with the demand growth, significantly contributed to the subsequent sharp decline of the FVE investment costs. This, along with subsidies (whether in the form of the Green bonus or high redemption prices of "clean" energy), led to an increase in the profitability of FVE investments which eventually became a lucrative business. The spinning spiral of the profitability of investment in FVE associated with the negative impact on the state budget was then slowed down by associated legislation by means of a gradual decline in subsidies until their eventual total abolition.

From the current perspective it seems that for various reasons the unsubsidized rooftop FVE of family houses and production halls are still profitable when consuming self-produced electricity; the yield on this investment is monetary savings on electricity that would otherwise have been drawn from the grid. The illustration of how the gradual reduction of investment costs and subsidies reflected in the profitability of investments in FVE, assessed according to **IRR**, is the subsequently solved problem.

3.1 Task assignment

The object of the task is the comparison of the profitability of investments in two rooftop FVEs installed on family houses in the CR. In both cases it deals with FVE of the same type and design, the only differences being the timing of FVE acquisition and the subsequent initialization of the operation, including the consequences that this entails. In the first case, the FVE was put into operation in January 2013 (FVE 2013), in the second case in January 2016 (FVE 2016). Both FVEs thus differ through acquisition costs of investment that were market-price at the time of their purchase and by the state subsidy policy. The task is to estimate the **IRR** of FVE investments for the time of their declared lifetime. It is expected that:

- All the electricity produced is consumed by the households (thereby, FVE 2016 generates savings of an unrealized offtake from the grid in the amount of CZK 5/kWh; FVE 2013 savings amount to CZK 5.49/kWh to which the Green bonus of CZK 2.28/kWh is added; the electricity prices that are constant over the lifetime of the projects, are the simplified assumption derived from average prices of 2012 and 2015).
- The system with the certified and brand components was declared by the manufacturer to have a minimum production lifetime of 25 years. The photovoltaic panels lose efficiency annually, which results in a loss of performance of about 1 %.
- The factor of annual utilization $k_r = W_r / (P_i \cdot h)$ oscillates in the CR below 12 % in the long term. The system with peak power $P_i = 1$ kWp produces on average W_r = 950-1,100 kWh of electricity per year (i.e. with the annual number of hours = 8,760 hours in a normal year, and 8,784 in a leap year).
- The initial investment in the acquisition, including the installation of FVE, ranged in 2013 from 35 to 59 thousand CZK per 1 kWp of installed capacity. For FVE 2016 it was necessary to invest between 29 and 37 thousand CZK. Rooftop FVE does not require any regular maintenance. Annual variable and fixed operating costs are both extremely low and can therefore be omitted.
- FVE systems are installed on south orientated sloping roofs with a 30° inclination.
- Since it deals with rooftop panels, for which the occurrence of a sudden power reduction PID (potential induced degradation) is rare, this is not taken into account.

3.2 Task solution

From the above stated assumptions and data the intervals $\mathbf{CF_0}$ = (–59, –47, –35) for FVE 2013, respectively $\mathbf{CF_0}$ = (–37, –33, –29) for FVE 2016 were compiled and significant points of intervals $\mathbf{CF} = (CF_l, CF, CF_r)$ for FVE 2013, respectively FVE 2016 were calculated that are valid for 100 % (initial) efficiency of the FV panels.

- For FVE 2013 **CF** = (7.38, 7.97, 8.55), where $CF_l = 7.38 = 0.95 \cdot (5.49 + 2.28)$ and
  $CF_r = 8.55 = 1.1 \cdot (5.49 + 2.28)$.
- For FVE 2016 **CF** = (4.75, 5.125, 5.5), where $CF_l = 4.75 = 0.95 \cdot 5$ and
  $CF_r = 5.5 = 1.1 \cdot 5$.

Significant points of intervals are given in thousands of CZK; since neither in the case of $\mathbf{CF_0}$ nor **CF** is there any relevant reason to suppose that the expected values CF_0 and CF are closer to one or the other extreme significant point, they were placed in the middle of the interval (CF_0 = (–59 + –35) / 2 = –47 and CF = (7.38 +8.55) / 2 = 7.97 for FVE 2013, respectively CF_0 = (–37 + –29) / 2 = –33 and CF = (4.75 + 5.5) / 2 = 5.125 for FVE 2016).

If FVE panels lose Δ % of their capacity annually (i.e. $CF_1 = CF$ and $CF_{i+1} = (1 - \Delta) \cdot CF_i$ for $i \geq 1$), then the cash flow generated by FVE in the i-th year of its operation $\mathbf{CF_i} = (1 - \Delta)^{i-1}$ **(·) CF** = $((1 - \Delta)^{i-1} \cdot CF_l, (1 - \Delta)^{i-1} \cdot CF, (1 - \Delta)^{i-1} \cdot CF_r)$. Substituting $CF_i = (1 - \Delta)^{i-1} \cdot CF$ in the middle equation of the system (6) we obtain

$$\sum_{i=1}^{n} (1 - \Delta)^{i-1} \cdot \frac{\mathrm{CF}}{(1 + \mathrm{IRR})^i} = -\mathrm{CF}_0. \tag{7}$$

Thence $\sum_{i=1}^{n}(1 - \Delta)^{i-1} \cdot CF / (1 + IRR)^i = \sum_{i=1}^{n}(1 - \Delta)^i \cdot (CF/(1 - \Delta))/(1 + IRR)^i = (CF / (1 - \Delta)) \cdot \sum_{i=1}^{n}(1 - \Delta)^i / (1 + IRR)^i = (CF / (1 - \Delta)) \cdot \sum_{i=1}^{n} 1/((1 + IRR)/(1 - \Delta))^i = (CF / (1 - \Delta)) \cdot \sum_{i=1}^{n} 1/(1 + (IRR + \Delta) / (1 - \Delta))^i = -CF_0$. Therefore

$$\sum_{i=1}^{n} 1/(1 + (\mathrm{IRR} + \Delta)/(1 - \Delta))^i = -\mathrm{CF}_0 \cdot (1 - \Delta)/\mathrm{CF} = \alpha\big((\mathrm{IRR} + \Delta)/(1 - \Delta), n\big). \tag{8}$$

The left side of equation (8) is an annuity factor $\alpha(r, n)$ of the n-year annuity (in our case n = 25) with the discount rate r = (IRR +Δ) / (1 – Δ), the middle expression is its desired value. Generally $\alpha(r, n)$ at fixed n and variable r is a simple projection α: (–1, ∞) → (0, ∞) in the form $\alpha(r, n) = \sum_{i=1}^{n} 1/(1 + r)^i$, $\alpha(0, n) = n$. Therefore, an inverse projection α^{-1}: (0, ∞) → (–1, ∞) exists in the form $r = \alpha^{-1}(y, n)$, where y is the required value $\alpha(r, n)$, $\alpha^{-1}(n, n) = 0$. The values of annuity factors for different r and n are tabulated and necessary values of functions α and α^{-1} can be easily obtained from tabulated values e.g. by linear interpolation.

If we search for IRR satisfying the condition $\alpha((IRR + \Delta) / (1 - \Delta), n) = -CF_0 \cdot (1 - \Delta) / CF$ declared by relation (8), we can proceed as follows: we search for a value $\alpha^{-1}(-CF_0 \cdot (1 - \Delta) / CF, n)$ in the table of annuity factors that should be equal to the value (IRR +Δ) / (1 – Δ). From there

$$\mathrm{IRR} = (1-\Delta)\cdot\alpha^{-1}(-\mathrm{CF}_0\cdot(1-\Delta)/\mathrm{CF}, n) - \Delta. \tag{9}$$

The analogous relations for the prediction of the extreme significant points $\mathrm{IRR_l}$, respectively $\mathrm{IRR_r}$ of the interval **IRR** are obtained by replacement of $-\mathrm{CF}_0$ and CF for $-\mathrm{CF_{0l}}$ and $\mathrm{CF_l}$, respectively for $-\mathrm{CF_{0r}}$ and $\mathrm{CF_r}$. In the case of $\Delta = 0.01$ and $n = 25$ we get:

$$\begin{aligned}\mathbf{IRR} &= (10.75\ \%, 15.61\ \%, 23.33\ \%) \text{ for FVE 2013,}\\ \mathbf{IRR} &= (11.14\ \%, 14.09\ \%, 17.72\ \%) \text{ for FVE 2016.}\end{aligned} \tag{10}$$

3.3 Discussion

The calculation results show that both analyzed profitability of investments in the same type of FVE launched at different times are comparable. This is mainly due to the merit of the subsidy policy, which by means of a configured funding system compensated differences between the actual investment costs and revenues from electricity production valid in the respective periods. The calculated lower limits $\mathrm{IRR_l}$ (10.75 % and 11.14 %) of intervals **IRR** are quite expressive: Even in the worst case the profitability of a particular investment is maintained above the level of 7 %, which, in the long term, corresponds to normal profitability in the electricity production sector (see [8] and [9]).

Even if the FVE 2013 had not been subsidized by the Green bonus (then its **CF** = (5.21, 5.62, 6.04), thus **IRR** = (6.37 %, 10.13 %, 15.93)), it would have been, albeit with the proviso mentioned below, acceptable. High values of **IRR** of the unsubsidized investment in FVE 2016 are due to a substantial decrease in acquisition costs of FVE between 2013 and 2016, whose influence suppressed the reversely acting tendencies caused by a reduction of subsidies for self-produced electricity. From the stated it is apparent that investments in rooftop FVEs, albeit reliant only on offtake savings from the grid, will pay off in the future. If the current status of acquisition costs is not accompanied by a corresponding reduction in electricity prices, or if the savings are not burdened by tax, the investments in unsubsidized rooftop FVEs can be considered as profitable opportunities.

A significant IRR point of the interval **IRR** is a measure of investment profitability according to which the investment profitability is primarily assessed within the standard approach to problem solving. The standard approach to decision-making according to the criteria IRR states: If the investment in FVE has IRR > 7 %, it is a profitable investment opportunity. Therefore, it should have been invested in FVE with IRR = 10.13 %. It would have been different in the case when the investor knew that **IRR** = (6.37 %, 10.13 %, 15.93), from which the IRR originates. The risk-averse investor (a rational investor) could have been unsettled by the occurrence of the possibility of $\mathrm{IRR_l}$ = 6.37 %, in which case he/she would not obtain normal profit in the power sector. If the normal profit is a zero economic profit, the border between the success (positive economic profit) and failure (economic loss) is the zero value of profitability. Given that the significant point $\mathrm{IRR_l}$ is the bottom limit of possible values, which the actual investment profitability will not be likely to exceed, the non-negative value $\mathrm{IRR_l}$ can be interpreted as a measure of investment safety. Analogously to the measure of safety of investment $\mathrm{IRR_l}$, a significant point $\mathrm{IRR_r}$ can be interpreted as a measure of hope for superior investment profitability.

The investment is called safe if its $\mathrm{IRR_l} > 0$. Since $\alpha^{-1}(y, n)$ is a monotonically decreasing function of the argument y, i.e. for $y_1 < y_2$ applies $\alpha^{-1}(y_1, n) > \alpha^{-1}(y_2, n)$, while $\alpha^{-1}(n, n) = 0$; for $y = -\mathrm{CF_{0l}}\cdot(1-\Delta)/\mathrm{CF_l} = n$ and thus for $-\mathrm{CF_{0l}}/\mathrm{CF_l} = n/(1-\Delta)$ from the equation (9) we get $\mathrm{IRR_l} = -\Delta$. In our case, for $n = 25$ and $\Delta = 0.01$ it applies that $\mathrm{IRR_l} = -1$ % at $-\mathrm{CF_{0l}}/\mathrm{CF_l} = 25/0.99 = 25.25$. The $\mathrm{IRR_l} = 0$ occurs (see equation (9)) at $\alpha^{-1}(-\mathrm{CF_{0l}}\cdot(1-\Delta)/\mathrm{CF_l}, n) = \Delta/(1-\Delta)$, i.e. at $\alpha(\Delta/(1-\Delta), n) = -\mathrm{CF_{0l}}\cdot(1-\Delta)/\mathrm{CF_l}$, from which $-\mathrm{CF_{0l}}/\mathrm{CF_l} = \alpha(\Delta/(1-\Delta), n)/(1-\Delta)$. In the case of $n = 25$ and $\Delta = 0.01$, $\mathrm{IRR_l} = 0$ at the value $-\mathrm{CF_{0l}}/\mathrm{CF_l} = 22$.

From the above the following conclusions result that are relevant to the choice between two investments in FVE:

- An investment in FVE with a production lifetime of $n = 25$ years and $\Delta = 0.01$ is safe (in terms of resistance to a potential financial loss), if its ratio $-\mathrm{CF_{0l}}/\mathrm{CF_l} < 22$.
- Of the two investments in FVE the investment with a lower ratio of the value $-\mathrm{CF_{0l}}/\mathrm{CF_l}$ is safer, with a lower value ratio $-\mathrm{CF_0}/\mathrm{CF}$ is more profitable and with a lower value ratio $-\mathrm{CF_{0r}}/\mathrm{CF_r}$ more hopeful (in terms of the possibility of premium profitability). (11)

4 Summary and conclusions

This contribution, in a certain sense, develops possibilities for the application of interval estimates for investment profitability with the purpose of improvement of methods of choice. These are the investments in production systems with robust linkages between inputs and outputs, working in conditions of uncertainty on the input side. The uncertainty refers to the state of vagueness, which does not provide a decision-maker with any guidance for the construction of rational estimates; however, possible ranges (intervals) of input values are partially known. The production system is expected to repeat its production process every year within its lifetime, while its effectiveness decreases by Δ % annually. The fixed component of the annual operating production costs is extremely low and is therefore neglected. A typical representative of such a production system is a solar photovoltaic power plant (FVE), presented in section 3.1.

The basic idea of the approach presented here for analyzing investment profitability is the thesis that in conditions of uncertainty there is a much greater chance to correctly predict the intervals at which the respective point values will occur than to correctly estimate these single values. Therefore, firstly the input variables were replaced in the standard NPV formula with the intervals of their possible values, which is represented in section 2 by the triplets of significant points. Thus, the relationship for NPV valid for point values was reformulated to the relationship for **NPV** valid in the language of intervals.

Within the herein established calculus of intervals the significant points of interval **NPV** were derived from the significant points of input intervals. From these the conditions were then established the significant points of **IRR** interval, the calculation of which is (due to the specific characteristics of the relevant cash flow) undemanding. The significant points **IRR** carry information not only about the level of profitability of the investment, but also about the level of its safety (in terms of resistance to a potential financial loss) and the degree of hope for the premium assessment. This approach offers more information about the benefits and drawbacks of the investment than a single point value of profitability derived through a standard method.

From the analysis of relationships between significant points of intervals $\mathbf{CF_0}$ and **CF** of input values on the one hand and significant points of interval **IRR** on the other, carried out in section 3.3, two important conclusions arose that are relevant to the selection between investments of the declared type. These conclusions are formulated in notification (11) and can be regarded as a major original contribution to this paper.

References

[1] Brealey, R. A., Myers, S. C., and Marcus, A. J.: *Fundamentals of Corporate Finance*. Mcgraw-Hill Education, New York, 2011.

[2] Danchev, S., Maniatis, G., and Tsakanikas, A.: Returns on investment in electricity producing photovoltaic systems under de-escalating feed-in tariffs: The case of Greece. *Renewable and Sustainable Energy Reviews* **14**,1 (2010), 500-505.

[3] Fiala, P.: Models for combined overbooking and capacity control in network revenue management. In: *Proceedings of the International Scientific Conference on Quantitative Methods in Economics - Multiple Criteria Decision* Making (Marian, R., ed.), Bratislava, 2012, 51-55.

[4] Hašek, R.: Systems of Computer Algebra and Dynamic Geometry as Tools of Mathematical Investigation. *International Journal for Technology in Mathematics Education* **20**(3) (2013), 103-108.

[5] Hašková, S.: Evaluation of Project Investments Based on Comprehensive Risk. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics (MME)* (Kocourek, A., and Vavrousek, M., eds.), Liberec, 2016, 260-264.

[6] Kahraman, C.: Capital budgeting techniques using discounted fuzzy cash flows. *Soft Computing for Risk Evaluation and Management*. Physica-Verlag HD (2001), 375-396.

[7] Kahraman, C., Murat, G., and Ulukan, Z.: Applications of fuzzy capital budgeting techniques. *Fuzzy Applications in Industrial Engineering*. Springer Berlin Heidelberg, 2006, 177-203.

[8] Kaldellis, J., K., Vlachou, D., S., and Korbakis, G.: Techno-economic evaluation of small hydro power plants in Greece: a complete sensitivity analysis. *Energy Policy* **33**,15 (2005), 1969-1985.

[9] Maroušek, J., Hašková, S., Zeman, R., Váchal, J., and Vaníčková, R.: Assessing the implications of EU subsidy policy on renewable energy in Czech Republic. *Clean Technologies and Environmental Policy* **17**,2 (2015), 549-554.

[10] Rodrigues, T. C., et al.: Modelling multicriteria value interactions with Reasoning Maps. *European Journal of Operational Research* **258**,3 (1017), 1054-1071.

[11] Roy, B., and Vincke, P.: Multicriteria analysis: survey and new directions. *European Journal of Operational Research* **8**,3 (1981), 207-218.

On Comparing Prediction Accuracy of Various EWMA Model Estimators

Radek Hendrych[1], Tomáš Cipra[2]

Abstract. The exponentially weighted moving average (EWMA) model is a particular modelling scheme advocated by RiskMetrics that is capable of predicting the current level of financial time series volatility. It is designed to track changes in conditional variance of financial returns by assigning exponentially decreasing weights to the observed past squared measurements. Recently, a recursive estimation technique suitable for this class of stochastic processes has been introduced and discussed. It represents a computationally attractive alternative to the already established non-recursive estimation strategies since it is effective in terms of memory storage, computational complexity and its ability to estimate and control the EWMA modelling scheme in real time. The aim of the paper is to investigate prediction accuracy of different EWMA model estimators. By analysing a set of eighteen very diverse world stock indices, this study has shown that the recursive estimation scheme can be recommended due to its advantageous properties if predicting the volatility; it is competitive to other approaches commonly used in financial practice.

Keywords: EWMA model, realized volatility, recursive estimation, RiskMetrics.

JEL classification: C51
AMS classification: 62M10

1 Introduction

The exponentially weighted moving average (EWMA) model is a modelling scheme preferred by the RiskMetrics methodology to evaluate conditional heteroscedasticity (see e.g. [9]). This concept is frequently linked to investigating financial time series, more specifically to monitoring volatility (i.e. the conditional standard deviation of financial returns). The EWMA model has been primarily developed as a simple alternative to the GARCH class models. The name of this concept originates from the fact that the conditional variance is an exponentially weighted sum of historical squared financial returns with the geometrically declining weights going backwards in time. Therefore, this model is easily capable of tracking changes in the conditional variance and volatility. Since its introduction in [9], it has been investigated from various perspectives with many successful empirical applications.

The only unknown parameter of the EWMA model that determines the geometrically declining weights is conventionally prescribed by experts or users. Alternatively, it can be estimated by employing common statistical identification procedures (e.g. the conditional maximum likelihood method). However, it is only exceptionally estimated recursively. On the other hand, it might be advantageous to apply a computationally effective technique that can estimate and control the model parameter (and consequently the model behaviour) in real time. Recently, several recursive estimation schemes suitable for this class of stochastic processes have been introduced (refer to e.g. [2], [4], [5] or [6]). These methods undoubtedly represent attractive alternatives to the conventional (i.e. non-recursive) estimation approaches due to their effectiveness in terms of memory storage and computational complexity. In particular, their adequacy has been demonstrated: In [4] the recursive estimation procedure suitable for the EWMA model has been originally proposed and thoroughly analysed by simulations. In [5] its prediction accuracy has been inspected by backtesting Value at Risk predictions.

The aim of this contribution is to investigate the prediction ability of different EWMA model estimators by using real financial data. Our attention is concentrated primarily on examining the recursive estimation techniques. It is motivated by recently published works (see [4], [5] and [6]), where the qualities of recursive estimation algorithms have been studied from other points of view as were briefly discussed above. The present paper should confirm the hypothesis that the recursive approach to estimation of the EWMA modelling parameter is not worse (at least) as the classical non-recursive estimation of this parameter, which is in practice usually based on the rolling-window technique. This conclusion seems to hold in addition to the fact that the recursive estimation approach respects various practical aspects including the computational simplicity and flexibility. The priorities of the recursive estimator are demonstrated here by means of an empirical case study of eighteen world stock indices.

[1]Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, hendrych@karlin.mff.cuni.cz.
[2]Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, cipra@karlin.mff.cuni.cz.

This paper is organized as follows. Section 2 recapitulates the EWMA modelling framework and its fundamental features. Section 3 briefly reviews the various estimation procedures investigated in this work: (*i*) the recursive estimator originally introduced in [4] and (*ii*) the rolling-window estimator based on the conditional log-likelihood criterion, which is commonly employed in practice. Section 4 describes an empirical study in which various estimation strategies leading to different volatility forecasts are compared. Section 5 concludes the paper.

2 EWMA modelling framework

The EWMA model of financial returns $\{y_t\}_{t\in\mathbb{Z}}$ is commonly defined as in [9]:

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = (1-\lambda) y_{t-1}^2 + \lambda \sigma_{t-1}^2, \tag{1}$$

where the only modelling parameter λ lies in the interval $(0,1)$, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables with zero mean and unit variance, and σ_t^2 is $\mathcal{F}_{t-1}$-measurable. Remind that $\mathcal{F}_t$ denotes the smallest σ-algebra with respect to which y_s is measurable for all $s \le t$, $s,t \in \mathbb{Z}$.

One can readily compute the first two conditional moments:

$$\mathbb{E}(y_t|\mathcal{F}_{t-1}) = 0, \quad \mathsf{var}(y_t|\mathcal{F}_{t-1}) = \mathbb{E}(y_t^2|\mathcal{F}_{t-1}) = \sigma_t^2, \quad \forall t. \tag{2}$$

Apparently, the positivity of the conditional variance σ_t^2 is ensured by construction of (1). In the given EWMA representation, more recent financial returns have greater weights on the conditional variance σ_t^2. The weights on financial returns decline geometrically going backwards in time by applying the weighting factors $(1-\lambda)$, $(1-\lambda)\lambda$, $(1-\lambda)\lambda^2, \ldots$. This is more evident from the equivalent representation of σ_t^2:

$$\sigma_t^2 = (1-\lambda) \cdot \sum_{j=0}^{\infty} \lambda^j y_{t-1-j}^2 \ \left[= (1-\lambda) y_{t-1}^2 + \lambda \sigma_{t-1}^2 \right]. \tag{3}$$

The one-step ahead prediction of σ_t^2 is expressed as:

$$\sigma_{t+1|t}^2 := \mathbb{E}(\sigma_{t+1}^2|\mathcal{F}_t) = (1-\lambda) y_t^2 + \lambda \sigma_t^2 = \sigma_{t+1}^2. \tag{4}$$

In the similar manner, the k-step ahead forecast of σ_t^2 (for $k \ge 1$) is given by :

$$\sigma_{t+k|t}^2 := \mathbb{E}(\sigma_{t+k}^2|\mathcal{F}_t) = \sigma_{t+1|t}^2. \tag{5}$$

3 Recursive estimation of EWMA model

In this section, we shall introduce the one-stage recursive estimation algorithm that can estimate the parameter λ of the EWMA process (1) in real time. In many instances, this approach may be truly advantageous. For example, it is possible to monitor or predict volatility on-line in the high-frequency financial data context. Recursive estimation methods are also effective in terms of memory storage and computational complexity since the current parameter estimates are evaluated using the previous estimates and actual measurements. Moreover, they can be used to detect structural model changes.

Applying general recursive prediction error method, see e.g. [7] or [8], one can derive the recursive scheme for estimating the parameter λ of the EWMA model (1). Principally, the negative conditional log-likelihood criterion corresponding to the EMWA process (refer also to (7)) is recursively minimized (when assuming normally distributed innovations ε_t). The resulting algorithm can be concisely formulated as follows (if one observes $y_1, y_2, \ldots$ one by one in time):

$$\begin{aligned}
\widehat{\lambda}_t &= \widehat{\lambda}_{t-1} + \frac{\widehat{p}_{t-1}(y_t^2 - \widehat{\sigma}_t^2)\widehat{\sigma}_t^{2'}}{\alpha_t(\widehat{\sigma}_t^2)^2 + (\widehat{\sigma}_t^{2'})^2 \widehat{p}_{t-1}}, \\
\widehat{p}_t &= \frac{1}{\alpha_t}\left\{ \widehat{p}_{t-1} - \frac{\widehat{p}_{t-1}^2 (\widehat{\sigma}_t^{2'})^2}{\alpha_t(\widehat{\sigma}_t^2)^2 + (\widehat{\sigma}_t^{2'})^2 \widehat{p}_{t-1}} \right\}, \\
\widehat{\sigma}_{t+1}^2 &= (1-\widehat{\lambda}_t) y_t^2 + \widehat{\lambda}_t \widehat{\sigma}_t^2, \\
\widehat{\sigma}_{t+1}^{2'} &= -y_t^2 + \widehat{\sigma}_t^2 + \widehat{\lambda}_t \widehat{\sigma}_t^{2'}, \ t \in \mathbb{N},
\end{aligned} \tag{6}$$

where $\widehat{\lambda}_t$ denotes the recursive estimate of the parameter λ at time t. We recommend initializing the foregoing procedure under the following conditions (analogously as it is outlined in [7]): (*i*) $\widehat{p}_0$ is a large positive number,

e.g. $\widehat{p}_0 = 10^5$, (*ii*) $\widehat{\lambda}_0$ should be taken from the interval $(0, 1)$, e.g. as 0.94 which is usually preferred for daily data, (*iii*) $\widehat{\sigma}_1^2$ is a positive number (e.g. the sample variance of several first measurements) and $\widehat{\sigma}_1^{2'}$ equals zero, (*iv*) $\{\alpha_t\}_{t\in\mathbb{N}}$ is a deterministic sequence of real positive numbers smaller or equal to one that either accelerates convergence or allows tracking parameter changes as it is discussed later in this section.

At each time t, it is necessary to check whether the current recursive estimate of λ belongs to the interval $(0, 1)$ before calculating other quantities in (6). If not, one should artificially replace the actual estimate by the previous one to avoid eventual specification problems. This simple projection rule ensures the positivity of conditional variance because $\widehat{\lambda}_0$ lies inside the interval. The sequence $\{\alpha_t\}_{t\in\mathbb{N}}$, the so-called *forgetting factor*, may be selected as follows: (*i*) α_t gradually grows to one as t goes to infinity, e.g. $\alpha_t = 0.99\alpha_{t-1} + 0.01$, $\alpha_0 = 0.95$, (*ii*) $\alpha_t \equiv \alpha$ for some $\alpha \in (0, 1)$ and all t, e.g. $\alpha = 0.990$. The first option corresponds to recursive estimation of the model (1) supposing time-invariant λ. The increasing forgetting factor significantly improves the convergence speed of the algorithm during the transient phase. The reason is that early information involved in several first measurements is somewhat misused (due to obvious initial uncertainties) and should therefore carry a lower weight when comparing to later measurements, which are processed in a better way. The second described case is associated with the eventuality that λ can vary over time. The constant forgetting factor less than one progressively reduces the influence of historical measurements, and thus enables to detect parameter changes. One may deduce (consult [7]) that measurements older than the sample of the length $T_0 = 1/(1 - \alpha)$ are involved in the recursive estimation with a weight that corresponds to $e^{-1} \times 100\%$ of that of the most recent measurement. T_0 is called a *memory time constant*. A suitable choice of $\alpha \in (0, 1)$ can be made from T_0 when assuming that the model parameter remains approximately constant over T_0 samples. See [8] for more details and other relevant references.

Theoretical properties of the suggested recursive estimation algorithm coincide with the conventional non-recursive case (as t goes to infinity), where the corresponding negative conditional log-likelihood criterion is minimized. Namely, convergence and asymptotic distributional properties are identical for a sufficiently large number of observations. Refer to [8] for the theoretical background of the prediction error method. The estimation scheme (6) can be further robustified in the similar manner as was reported in [4].

The introduced recursive estimation technique (6) might be an appealing alternative to the rolling-window estimation approach. The latter approach is based on repeating minimization of the negative conditional log-likelihood function, which corresponds to the model (1) when assuming normally distributed innovations ε_t. It can be expressed as follows:

$$\widehat{\lambda}_t^M = \arg\min_{\lambda\in(0,1)} \sum_{\tau=t-M+1}^{t} \left[\frac{y_\tau^2}{\sigma_\tau^2} + \log(\sigma_\tau^2)\right], \; t \geq M, \tag{7}$$

where $M \in \mathbb{N}$ denotes the rolling-window width. At each time the minimum (7) is repeatedly calculated and $\widehat{\sigma}_{t+1}^2$ is evaluated using the most recent estimate $\widehat{\lambda}_t^M$ and M consecutive observations. The estimation is initialized similarly as above. It is obvious that the estimation can start only after observing at least M financial returns. This scheme or its alternatives are frequently applied in practice. However, such an estimator is computationally very complex since the optimization task is obviously solved by an iterative procedure based on M consecutive measurements at each time. Alternatively, one could reestimate the unknown parameter less frequently.

4 Empirical study: Volatility forecast comparison

4.1 Layout of study

In this section, we shall examine the performance of the suggested recursive estimator of the EWMA model (6) when analysing an empirical dataset. The proposed estimation procedure is compared with the rolling-window estimation of the EWMA model (7) and the EWMA model with a fixed value of the parameter λ. To be more precise, we will study daily observed close-to-close financial returns of eighteen stock indices (see Section 4.2 for more details). For each day t in the forecasting sample, we have estimated (calibrated) the EWMA model by employing a particular estimation strategy using data at or before t (i.e. conditioning information has varied in time). Afterwards, we have applied these fitted models to predict volatility using the relation (5) for the one-day prediction horizon. This process has generated a sequence of daily forecasts for each estimation approach and each analysed time series of financial returns. These forecast paths are then further studied.

The measure of volatility predictive accuracy is based on the average forecast loss achieved by the EWMA model estimated by a particular estimation technique. The estimation procedure that provides a smaller average loss is more accurate and therefore preferred. One can implement a variety of loss functions with distinct statistical properties. In [10], there was identified a class of robust loss functions that asymptotically generate the same ranking of estimation approaches regardless of the ex-post volatility proxy being used (as long as the proxy is unbiased and the minimal regularity conditions are met). Note that the volatility is not directly observed on markets; therefore, forecasts must be compared with an ex-post volatility proxy such as realized volatility or squared returns.

Accepting these arguments and respecting the discussion provided in [1], we shall prefer the quasi-likelihood (QL) forecast loss function given as:

$$QL\left(\widehat{\sigma}^2_{t+k,P}, \widehat{\sigma}^2_{t+k|t}\right) = \frac{\widehat{\sigma}^2_{t+k,P}}{\widehat{\sigma}^2_{t+k|t}} - \log\left(\frac{\widehat{\sigma}^2_{t+k,P}}{\widehat{\sigma}^2_{t+k|t}}\right) - 1,\ k \geq 1, \tag{8}$$

where $\widehat{\sigma}^2_{t+k,P}$ is an unbiased ex-post volatility proxy (such as realized volatility or squared financial returns) and $\widehat{\sigma}^2_{t+k|t}$ is a volatility forecast given by (5) when applying a corresponding estimation procedure.

The QL criterion is preferred for two main reasons. Firstly, the loss series defined as (8) is i.i.d. under the null hypothesis that the forecasting model is correctly specified (consult [1]). Secondly, assume that the volatility proxy $\widehat{\sigma}^2_{t+k,P}$ can be expressed as $\widehat{\sigma}^2_{t+k,P} = h_{0,t+k}\zeta_{t+k}$, where $h_{0,t+k}$ is the latent true variance and ζ_{t+k} is i.i.d. measurement error with unit expected value and variance $\delta^2 \in (0, \infty)$. The conditional expected value of QL is:

$$\mathbb{E}\left[QL\left(\widehat{\sigma}^2_{t+k,P}, \widehat{\sigma}^2_{t+k|t}\right)|\mathcal{F}_t, h_{0,t+k}\right] \approx QL\left(h_{0,t+k}, \widehat{\sigma}^2_{t+k|t}\right) + \frac{1}{2}\delta^2, \tag{9}$$

where one has employed a standard Taylor expansion for moments of a random variable. Hence, it is clear that the bias of QL is independent of the volatility level. It makes easier to compare losses across different volatility regimes.

Stock Index	***M1***	***M2***	***M3***	***M4***	***M5***	***M6***
S&P 500	0.26475 [4]	0.26290 [3]	**0.26027 [1]**	0.26255 [2]	0.26975 [6]	0.26731 [5]
FTSE 100	0.19990 [3]	**0.19557 [1]**	0.20829 [5]	0.19622 [2]	0.22737 [6]	0.20310 [4]
Nikkei 225	0.25589 [3]	**0.24879 [1]**	0.32554 [6]	0.25494 [2]	0.25733 [4]	0.26149 [5]
DAX	0.20288 [4]	**0.20003 [1]**	0.20127 [3]	0.20051 [2]	0.22558 [6]	0.21507 [5]
Russel 2000	0.37078 [6]	0.36225 [4]	0.35252 [2]	0.36279 [5]	0.36064 [3]	**0.35216 [1]**
All Ordinaries	0.27152 [5]	0.27146 [4]	**0.26551 [1]**	0.26690 [2]	0.28043 [6]	0.26810 [3]
DJIA	0.28149 [3]	**0.27910 [1]**	0.28541 [4]	0.28009 [2]	0.31266 [6]	0.29241 [5]
Nasdaq 100	0.22863 [5]	0.23567 [6]	0.22453 [3]	0.22782 [4]	0.22444 [2]	**0.22205 [1]**
CAC 40	0.18623 [2]	**0.18363 [1]**	0.19379 [4]	0.18735 [3]	0.21518 [6]	0.19718 [5]
Hang Seng	**0.18351 [1]**	0.18849 [5]	0.18422 [2]	0.18634 [3]	0.19356 [6]	0.18717 [4]
KOSPI Composite Index	**0.18135 [1]**	0.18236 [2]	0.18825 [4]	0.18518 [3]	0.20335 [6]	0.18838 [5]
AEX Index	0.20313 [4]	**0.19557 [1]**	0.19718 [2]	0.19865 [3]	0.22345 [6]	0.20605 [5]
Swiss Market Index	0.18689 [5]	0.17984 [2]	0.18007 [3]	**0.17880 [1]**	0.20068 [6]	0.18398 [4]
IBEX 35	0.20076 [2]	**0.19905 [1]**	0.21000 [5]	0.20077 [3]	0.22436 [6]	0.20960 [4]
S&P CNX Nifty	0.28060 [3]	**0.27237 [1]**	0.28133 [4]	0.27272 [2]	0.30040 [6]	0.28275 [5]
IPC Mexico	0.47831 [5]	0.48874 [6]	**0.46312 [1]**	0.47444 [4]	0.47293 [3]	0.47269 [2]
Bovespa Index	**0.20803 [1]**	0.21892 [6]	0.20824 [2]	0.21341 [5]	0.21275 [4]	0.20832 [3]
Euro STOXX 50	0.23533 [4]	**0.22966 [1]**	0.23173 [2]	0.23355 [3]	0.24621 [6]	0.23746 [5]
Sums of ranks	61	**47**	54	51	94	71
Averages of ranks	3.39	**2.61**	3.00	2.83	5.22	3.94

Table 1 Average *QL* losses for one-day volatility predictions (ranks in square brackets)

4.2 Numerical results

In this section, we will thoroughly investigate six different estimation variants suitable for the EWMA model. Particularly, we shall evaluate the following estimation approaches: (*M1*) the parameter λ is fixed as it is common for daily data, i.e. $\lambda = 0.94$ (see [9]), (*M2*) the parameter λ is estimated recursively by employing (6) with the forgetting factor given as $\alpha_t = 0.99\alpha_{t-1} + 0.01$, $\alpha_0 = 0.95$, (*M3*) it is similar to (*M2*) but with the forgetting factor $\alpha_t = 0.990$ (i.e. $T_0 = 100$), (*M4*) it is similar to (*M2*) but with the forgetting factor $\alpha_t = 0.996$ (i.e. $T_0 = 250$), (*M5*) the parameter λ is estimated by the rolling-window estimator (7) with $M = 100$, (*M6*) it is similar to (*M5*) but with $M = 250$.

The present empirical study compares various volatility forecast strategies by studying a portfolio of eighteen stock indices. The examined dataset contains daily close-to-close financial returns and a corresponding sequence of daily realized volatilities calculated using five minutes intraday data as an ex-post volatility proxy. In particular, the realized volatility is a high frequency, non-parametric based estimator of the variation of the price path of an asset during the times at which the asset trades frequently on an exchange. It ignores the variation of prices overnight and sometimes the variation in the first few minutes of the trading day when recorded prices may contain large errors. To be more precise, the data and its description can be found in [3]. The analysed dataset is composed of more than 4400 measurements for each stock index; the daily close-to-close financial returns were observed between 3 January 2000 and 13 January 2017. All calculations were performed using software `R`.

Table 1 displays average quasi-likelihood (*QL*) losses (8) for all estimation variants (*M1*) - (*M6*) and all considered stock indices. We have additionally assigned ranks to each estimation alternative (*M1*) - (*M6*) for all examined

indices. Afterwards, we have calculated sums of the ranks and their averages in order to evaluate the most accurate EWMA model estimation strategy. The outputs for the one-day prediction horizon might be summarized as follows: The three recursively estimated conditional variances (*M2*) - (*M4*) has achieved the best accuracy. They are at least competitive to the remaining estimation strategies. The estimation strategy (*M5*) seems to be less satisfactory (regardless of its high computational complexity). Moreover, Figure 1 displays one-day forecasts of the EWMA conditional variances (*M1*) - (*M6*) (solid lines) together with the realized variances (crosses) for the S&P 500 stock index. One can identify similar volatility trends across all the graphs, especially the peaks in the financial crisis period 2008-2009. Analogical figures could be plotted for any of other assets.

The comprehensive conclusion delivered by this empirical analysis confirms that suggested recursive estimation algorithm (6) seems to be competitive to others; it has outperformed the rolling-window estimator (7), which has, moreover, much higher computational complexity in comparison with (6).

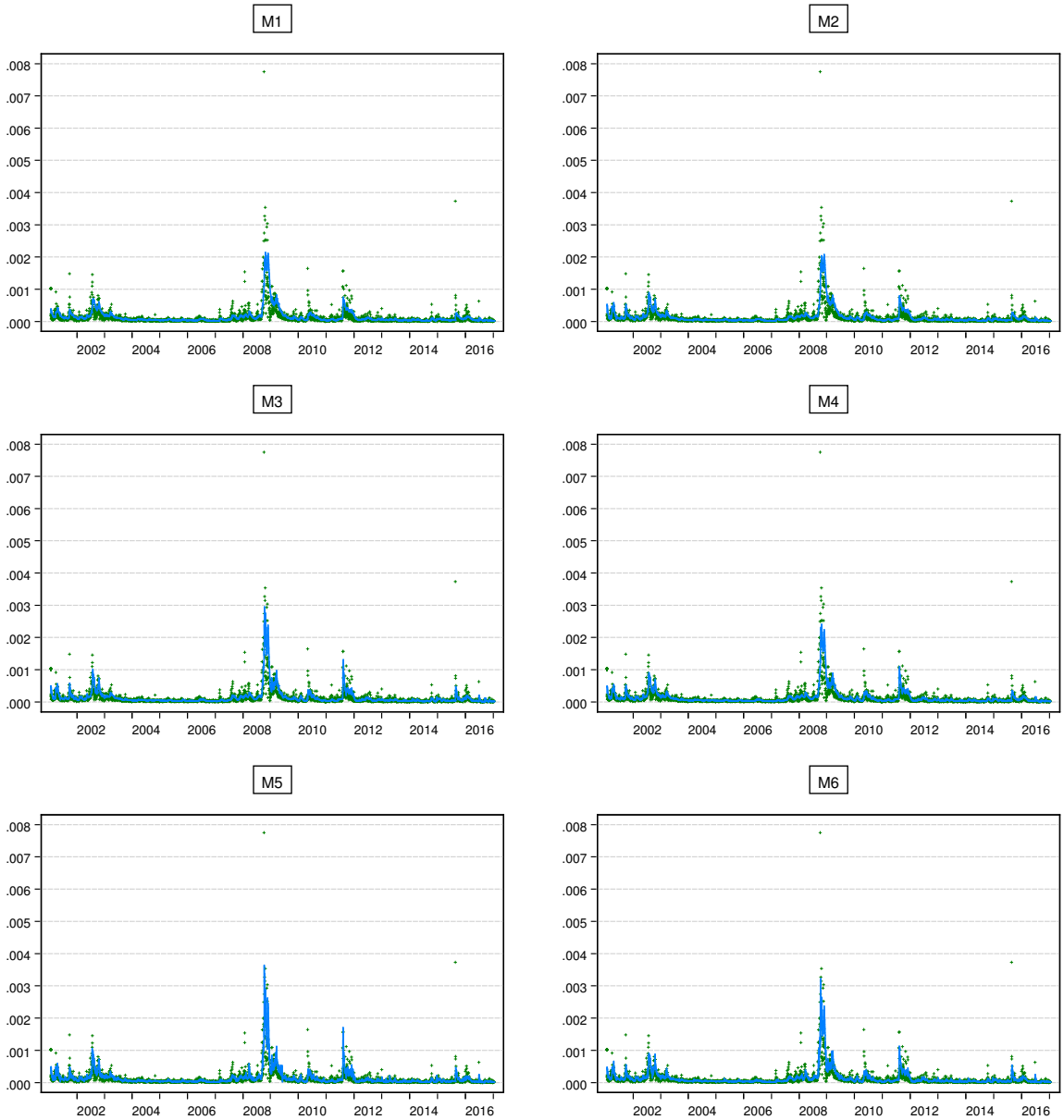


Figure 1 Daily conditional variance forecasts (solid lines) with realized variances (crosses) for S&P 500

5 Conclusion

The volatility estimation and prediction play an important role in modern finance, especially, if one deals with risk evaluation. The present paper examined various aspects of the recursive estimation of the EWMA model including technical instructions for its application. It justified that the recursive estimation scheme can be recommended in financial practice. In particular, the empirical study reviewed some facts on volatility predictions in finance. By investigating a set of eighteen very diverse stock indices, the analysis has shown that the recursive estimation technique can be recommended due to its advantageous properties if predicting the volatility; the recursive estimation algorithm is competitive to other approaches commonly used in financial practice.

Acknowledgements

This research was supported by the grant P403/17/676S provided by the Grant Agency of the Czech Republic.

References

[1] Brownlees, C., Engle, R., and Kelly, B.: A practical guide to volatility forecasting through calm and storm. *The Journal of Risk* **14** (2011), 3–22.

[2] Gerencsér, L. and Orlovits, Z.: Real time estimation of stochastic volatility process. *Annals of Operations Research* **200** (2012), 223–246.

[3] Heber, G., Lunde, A., Shephard, N., and Sheppard, K.: *Oxford-Man Institute's realized library, version 0.2*. Oxford-Man Institute, Oxford University, Oxford, 2009. `http://realized.oxford-man.ox.ac.uk/`, accessed 19 January 2017.

[4] Hendrych, R.: Robustified on-line estimation of the EWMA models: Simulations and applications. In: *Proceedings of the 33rd International Conference Mathematical Methods in Economics* (D. Martinčák, J. Ircingová, and P. Janeček, eds.), University of West Bohemia, Plzeň, 2015, 237–242.

[5] Hendrych, R. and Cipra, T.: On comparing various EWMA model estimators: Value at Risk perspective. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics* (A. Kocourek and M. Vavroušek, eds.), Technical University of Liberec, Liberec, 2016, 265–270.

[6] Hendrych, R. and Cipra, T.: Self-weighted recursive estimation of GARCH models, *Communications in Statistics - Simulation and Computation* (2015), in press (doi: 10.1080/03610918.2015.1053924).

[7] Ljung, L.: *System Identification: Theory for the User*. Prentice Hall, Upper Saddle River (NJ), 1999.

[8] Ljung, L. and Söderström, T. S.: *Theory and Practice of Recursive Identification*. MIT Press, Cambridge (MA), 1983.

[9] Morgan, J.: *RiskMetrics - Technical document*. (4th ed.). Morgan Guaranty Trust Company, New York, 1996.

[10] Patton, A. J.: Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* **160** (2011), 246–256.

Robust optimization approach in transportation problem

Robert Hlavatý[1], Helena Brožová[2]

Abstract. In this paper we get back to the classical transportation problem of Hitchcock and Koopmans while considering some parameters within its mathematical programming formulation to be uncertain. It is a common practice that the both supply and demand of some actors can be indeterminate as well as the evaluation of routes between them in terms of distance and time. There are ways to deal with uncertainty in the linear optimization models such as stochastic optimization or interval programming. These approaches bring also computational difficulties since their practical application is quite demanding. We show how this situation can be handled using robust optimization approach towards modelling of uncertainty. The robust approach is a tool that seeks a robust-optimal solution and it allows, in its extended form, to specify the deviations from deterministic values quite precisely while the entire mathematical model remains simple which is a considerable computational advantage. The way of transformation of transportation problem into its robust counterpart is described in detail and accompanied by example of different scenarios in the end.

Keywords: Transportation problem, robust optimization, uncertainty

JEL Classification: C61
AMS Classification: 90C05, 90C08, 90C11

1 Introduction

A transportation problem is one of the first and also best known applications of linear programming (LP) problems. Originally formulated by Monge (1781) and developed by Hitchcock and Koopmans in the times of WWII, it was shown how one is able to reformulate a problem of transportation from several sources to a number of target locations as a LP problem. It should be mentioned that it was not only Hitchcock and Koopmans dealing with this matter during the forties but it is often considered as most important. The very first overview of the problem can be found in the work of Hitchcock (1941). The model itself is considered deterministic without any assumptions of uncertainty.

Parameter uncertainty in mathematical programming has been treated in many ways since Dantzig's pioneering work (we refer to Dantzig 1955). The Robust optimization or robust control is hardly a new topic in the field of mathematical programming. Soyster (1973) was one of the first researchers to pay attention to robust solutions in his paper. In the recent years however, the robust optimization has become a focus of many researchers as it happens to be somehow attractive for its not-so-stochastic nature. The robust optimization is deterministic and set-based rather than stochastic (Bertsimas, Brown and Caramanis, 2008).

From the recent years we would like to especially mention an extensive work of Ben-Tal, El Ghaoui and Nemirovski (2009) describing the issues of Robust optimization as a whole. A reader might want to inspect earlier work of Ben-Tal and Nemirovski (1999) that is written in more concise fashion and is dedicated solely to linear programming (LP) optimization problems. The earlier approaches like the one of Soyster's (1973) tend to reach the solutions that are over-conservative. Ben-Tal and Nemirovski (1999) propose an approach to overcome the solution conservativeness by using ellipsoidal uncertainty sets. This results into a requirement of conic quadratic programming when the robust solution is searched for and it becomes rather complex and not applicable to discrete optimization problems. Bertsimas and Sim (2003) propose a different approach to control the level of conservatism in the solution in a manner that leads to a linear optimization model that can even be applied to discrete optimization problems. A number of authors have adapted the approach of Bertsimas and Sim (2003) with the slight changes as Palma and Nelson (2009) did, for instance.

Recently, a new extension of robust optimization approach was developed called multi-band robust optimization. It utilizes the concept of standard robust optimization and brings a lot of new modelling possibilities, especially in the matters of uncertainty modelling. Using the multi-band approach, one is able to describe an uncertain modelling situation more precisely and thus, receive more precise results after all. The multi-band approach combines standard robust optimization approach with interval programming and allows a decision maker to set the uncertainties in some particular ranges while it is still kept a linear programming problem. This is very

[1] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, hlavaty@pef.czu.cz
[2] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, brozova@pef.czu.cz

important when working with the large scale models as it is well known that the linear programming models are mostly computationally tractable and the solution can be found within a reasonable time.

The multi-band approach itself has been recently under development and continuous improvements. The credit goes to Büsing and D'Andreagiovanni (2012) who can be considered current pioneers of multi-band optimization approach. One year later, their research results in a general review of the multi-band optimization described in Büsing and D'Andreagiovanni (2013) and outlines the entire theoretical foundations of the approach. It is followed by other expansions like Büsing and D'Andreagiovanni (2014) where the same authors present a new theoretical framework of the problem.

The paper explores the applicability of robust approach methodologies into transportation problem. There have been several scientific papers dealing with the robust optimization in practice but mostly not on this matter. Regarding the transportation problems where the robust approach was utilized there are some recent works. Maggioni, Potra and Bertocchi (2014) show a robust multi-stage transportation problem case study in cement industry, Gabrel et al (2015) show a robust location transportation problem under uncertain demands and finally, there is an important work of Ben-Tal et al (2011) dealing with the emergency logistics planning. In this paper, we would like to bring a simple guide for a reader who wants to get familiar with the robust programming in practice. It will be shown how to reformulate a very well-known transportation problem into its robust counterpart.

2 Materials and methods

In this section, a way of transformation of a generic linear optimization model to its robust counterpart will be described, followed by mathematical description of a general transportation problem and building a robust counterpart for the transportation problem.

2.1 Robust programming introduction

In this paragraph, it is described how a linear optimization model can be transformed into its robust counterpart, that is, a modified program that seeks robust optimal solution. Such solution is resistant against any deviations that might occur anywhere in the original deterministic model. For the description we utilize the approach of "Γ-robustness" of Bertsimas and Sim (2003). We assume a generic linear optimization model:

$$\begin{aligned} & max \sum_{j=1}^{n} c_j x_j \\ & \text{s.t.} \\ & \sum_{j=1}^{n} a_{ij} x_j \leq b_i \,, i = 1, \dots, m \\ & x_j \geq 0 \,, j = 1, \dots, n \end{aligned} \tag{1}$$

When the presence of uncertainty is considered we usually assume that the problem coefficients a_{ij}, b_j, c_j (or at least some of them) are not precisely defined. This leads to a new problem:

$$\begin{aligned} & max \sum_{j=1}^{n} \left(c_j + \delta_j^c\right) x_j \\ & \text{s.t.} \\ & \sum_{j=1}^{n} \left(a_{ij} + \delta_{ij}^a\right) x_j \leq b_i + \delta_i^b \,, i = 1, \dots, m \\ & x_j \geq 0 \,, j = 1, \dots, n \end{aligned} \tag{2}$$

The (2) represents the reformulation of (1) with uncertain coefficients. The uncertainty is expressed using deviations for any coefficients if need be. We assume any deviation δ to be any real nonzero number. It was illustrated by Ben-Tal, El Ghaoui and Nemirovski (2009) that even a slight change in the original coefficient value may affect the optimal solution adversely. In some cases it may even become infeasible while this is caused by a small deviation. For simplicity, let us consider the uncertainty to be present within the set of coefficients a_{ij} from now on. Before a construction of the robust counterpart of (1), Bertsimas and Sim (2003) propose 4 assumptions that must hold:

1) For each coefficient a_{ij}, one is able to define its deterministic (expected, usual,…) value and its maximum deviation δ_{ij}^a from the deterministic value.

2) The deterministic value a_{ij} then belongs to the symmetric interval $\left[a_{ij} - \delta_{ij}^{a}, a_{ij} + \delta_{ij}^{a}\right]$
3) The uncertain coefficients are stochastically independent random coefficients, each of the with its own deviation range
4) For each constraint i, one is able to define a maximum number of coefficients Γ_i that will deviate simultaneously from its deterministic value in the constraint i.

Then, a robust model can be constructed, based upon deviations and it is possible to reach the final form of this robust counterpart of (1):

$$\begin{gathered} max \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \\ \sum_{j=1}^{n} a_{ij} x_j + \Gamma_i z_i + \sum_{j \in \mathcal{U}_i} p_{ij} \leq b_i \,, i = 1, \dots, m \\ z_i + p_{ij} \geq \delta_{ij}^{a} x_j \,, i = 1, \dots, m \,, \forall j \in \mathcal{U}_i \\ z_i \geq 0 \,, i = 1, \dots, m \\ p_{ij} \geq 0 \,, i = 1, \dots, m \,, \forall j \in \mathcal{U}_i \\ x_j \geq 0 \,, j = 1, \dots, n \end{gathered} \tag{3}$$

where the parameter $\Gamma_i, 0 \leq \Gamma_i \leq |\mathcal{U}_i|, i = 1, \dots, m$, controls the protection against uncertainty in the constraint i, p_{ij} is an auxiliary variable for each a_{ij} that is considered uncertain, z_i is another auxiliary variable merely preserving a relationship between the first and second constraint and $\mathcal{U}_i$ indicates a set of indices j of those a_{ij} for which the deviation is actually considered. The creation of (3) is based on exploiting the properties of primal and dual versions of the original problem (1).

2.2 LP formulation of Transportation problem

We consider a general transportation problem of an unidentified homogenous resource from m suppliers to n target locations while minimizing the total transportation cost. Further, we consider these suppliers to provide nonzero amounts $a_1, a_2, \dots, a_m$ and the target locations demanding amounts $b_1, b_2, \dots, b_m$ respectively. W.l.o.g. we suppose that the problem is balanced in terms of overall supply equal to overall demand. Each route from any suppliers to any target location is evaluated with a cost coefficient $c_{ij}, i = 1, \dots m, j = 1, \dots, n$. The optimization model with is formulated as:

$$\begin{gathered} min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\ \text{s.t.} \\ \sum_{j=1}^{n} x_{ij} = a_i \,, i = 1, \dots, m \\ \sum_{i=1}^{m} x_{ij} = b_j \,, j = 1, \dots, n \\ x_{ij} \geq 0, \forall i, j \end{gathered} \tag{4}$$

where the decision variables x_{ij} represent the amount to be transported from supplier i to target location j.

3 Results

With the transportation problem and robust counterpart of LP defined it is now possible to join them together in order to create formulation of robust transportation problem. It was shown in (2) that three categories of coefficients can be considered to be uncertain: left-hand side coefficients, right-hand side coefficients and cost coefficients. One can notice that considering a change in left-hand side coefficients in the transportation problem is not an option in this case since here we have all of these coefficients equal to one necessarily. The right-hand side coefficients representing the capacities could be theoretically considered uncertain, however, in practice we assume that we are certain what these capacities would be. On the other hand, the transportation cost is the parameter that can show some uncertain behavior. The transportation cost on a particular route can be affected by a couple of undesirable factors like detours, traffic jams etc. that would only raise the costs.

3.1 Formulation of robust counterpart for transportation problem

Due to practical reasons, we will consider that the uncertainty will be laid on objective function coefficients c_{ij} only. In the description of the robust counterpart of LP (3) it was only shown in general how to construct robust counterpart for the case of deviations appearing in left-hand side coefficients. Let us show that this is not an obstacle in case the uncertainty is considered in objective function coefficients. We consider the following objective function (of a general transportation problem) with the deviations in c_{ij}:

$$min \sum_{i=1}^{m} \sum_{j=1}^{n} (c_{ij} + \delta_{ij}^{c}) x_{ij} \tag{5}$$

This expression (5) is equivalent to the following:

$$\begin{gathered} \min \Phi \\ \text{s.t.} \\ \sum_{i=1}^{m} \sum_{j=1}^{n} (c_{ij} + \delta_{ij}^{c}) x_{ij} \leq \Phi \end{gathered} \tag{6}$$

The coefficient c_{ij} in (6) can now be treated as left-hand side coefficient and the same approach for including the uncertainty can be applied according to (3). Similarly, it is possible to turn right-hand side coefficients into left-hand side coefficient and treat them so, if needed.

It is now possible to define the robust counterpart for a general transportation problem with the uncertain coefficients in objective function:

$$\begin{gathered} \min \Phi \\ \text{s.t.} \\ \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} + \Gamma z + \sum_{(i,j) \in \mathcal{U}} p_{ij} \leq \Phi \\ z + p_{ij} \geq \delta_{ij}^{c} x_{ij} \, , \forall (i,j) \in \mathcal{U} \\ \sum_{j=1}^{n} x_{ij} = a_i \, , i = 1, \dots, m \\ \sum_{i=1}^{m} x_{ij} = b_j \, , j = 1, \dots, n \\ z \geq 0 \\ p_{ij} \geq 0 \, , \forall (i,j) \in \mathcal{U} \\ x_{ij} \geq 0 \, , i = 1, \dots, m \, , j = 1, \dots, n \end{gathered} \tag{7}$$

where the parameter $\Gamma, 0 \leq \Gamma \leq |\mathcal{U}|$, controls the protection against uncertainty in the cost coefficients $c_{ij}, (i,j) \in \mathcal{U}$ of transportation problem. Variables p_{ij} and z have the equivalent meaning as in (3).

3.2 Practical example

To illustrate how it is possible to implement the aspects of uncertainty into the transportation problem and how the change of initial parameters influences the objective value, we will assume the following transportation problem with parameters in the cost matrix C, suppliers' capacities $\boldsymbol{a}$ and consumer demands $\boldsymbol{b}$:

$$C = \begin{pmatrix} 2 & 3 \\ 8 & 1 \\ 3 & 6 \end{pmatrix} \quad \boldsymbol{a} = \begin{pmatrix} 50 \\ 100 \\ 120 \end{pmatrix} \quad \boldsymbol{b} = \begin{pmatrix} 160 \\ 110 \end{pmatrix} \tag{8}$$

Solving the program (8) we receive an optimal solution vector of the problem $\boldsymbol{x}^* = (40,10,0,100,120,0)$ with the optimal objective value of 590. Now, let as assume that the following maximum deviations matrix Δ corresponding to assumed deviations for the coefficients in C:

$$\Delta = \begin{pmatrix} 1.2 & 4 \\ 4.8 & 0.6 \\ 4 & 3.6 \end{pmatrix} \tag{9}$$

where, for instance, $\delta_{11} = 1.2$ expresses that the cost coefficient $c_{11} = 2$ can possibly worsen (increase) by 1.2 units.

We will inspect six scenarios of robust solutions in dependence on parameter Γ that represents a protection against uncertainty, i.e. the number of coefficients c_{ij} that will deviate by δ_{ij} in the objective function. For example, setting $\Gamma = 2$ expresses that at most two coefficients c_{ij} will deviate from its deterministic value while it is not given which particular c_{ij} it would be. Obviously, the higher value of Γ will cause more robust solutions because more deviations will be assumed and that will also cause an increase in objective values. In the following table, the comparison of the results solved as (7) is provided:

Γ value		x_{11}	x_{12}	x_{21}	x_{22}	x_{31}	x_{32}	Objective value
$\Gamma = 0$	Solution vector:	40	10	0	100	120	0	570
	Real costs:	2	3	8	1	3	6	
$\Gamma = 1$	Solution vector:	50	0	0	100	110	10	1030
	Real costs:	2	3	8	1	7	6	
$\Gamma = 2$	Solution vector:	50	0	0	100	110	10	1090
	Real costs:	2	3	8	1.6	7	6	
$\Gamma = 3$	Solution vector:	50	0	0	100	110	10	1150
	Real costs:	3.2	3	8	1.6	7	6	
$\Gamma = 4$	Solution vector:	45.26	4.74	0	100	114.74	5.26	1172.74
	Real costs:	3.2	7	8	1.6	7	6	
$\Gamma = 5$	Solution vector:	50	0	0	100	110	10	1186
	Real costs:	3.2	3	8	1.6	7	9.6	
$\Gamma = 6$	Solution vector:	50	0	0	100	110	10	1186
	Real costs:	3.2	3	8	1.6	7	9.6	

Table 1 Different robust solutions

Table 1 shows how the solution grows more robust with increasing Γ. In case of $\Gamma = 0$, we suppose that no cost coefficients will deviate and thus we obtain the solution identical to deterministic solution. On the contrary, the worst case of $\Gamma = 6$ means that all cost coefficients can deviate to their worse values and the minimal guaranteed objective is then more than doubled compared to deterministic situation. The grey values in the table indicates those cost coefficients that actually deviated in the solution for each scenario. The scenario for $\Gamma = 4$ shows non-integer values. This is caused by the transformation of the original problem into robust counterpart which contains some new constraints and it need not satisfy the integrality of the results. The last two scenarios provide the same results even though the degree of uncertainty is different. This can happen since the interpretation of Γ value is "at most Γ coefficients will deviate". But even if all coefficients deviated to their maximum, in the last scenario for instance, then it would have no influence on the objective value since the variables x_{12}, x_{21} are not chosen into solution (the optimal value is 0).

4 Conclusion

The modelling approach of Γ-robustness was applied to the specific case of optimization area. In the transportation problem in practice uncertain conditions can be naturally assumed, especially due to instable traffic conditions on the routes between suppliers and target locations. The robust solutions show the worst-case scenarios for assumed deviations with the minimal guaranteed objective function. This can be useful for a decision maker to know in advance what would be the expected total transportation cost if some unexpected events occur on the transportation routes. This approach is much less demanding contrary to evaluation of all possible programs with the different parameters. Still, the robust optimization model is completely linear and thus easily computationally tractable.

Due to the nature of the robust model, some values of the solution vector need not to be integer, as it is expected in case of classical deterministic transportation model. If only integer values are desired, then the integrality conditions can be included for all x_{ij} in the robust model as it is also valid for integer linear programs. Further, we would like to explain the multi-band robustness that allows to model possible parameter deviations with better precision.

References

[1] Ben-Tal, A., El Ghaoui A.L. and Nemirovski A.: *Robust optimization*. Princeton University Press, 2009.

[2] Ben-Tal, A. and Nemirovski, A.: Robust solutions of uncertain linear programs. *Operations Research Letters* **25** (1999), 1-13.

[3] Ben-Tal, A., Chung, B.D., Mandala, S.R. and Yao, T.: Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains. *Transportation Research Part B: Methodological* **45** (2011), 1777-1189.

[4] Bertsimas, D., Brown, D.B. and Caramanis, C.: *Theory and applications of robust optimization*. Working paper, Massachusetts Institute of Technology, Cambridge, MA, 2008.

[5] Bertsimas, D. and Sim, M.: Robust discrete optimization and network flows. *Mathematical Programming* **98** (2003), 49-71.

[6] Büsing, C. and D'Andreagiovanni, F.: Newe results under multi-band uncertainty in robust optimization. In: *Proc. Of the 11th symposium on Experimental Algorithms – SEA 2012*, Heidelberg, 2012, 63-74.

[7] Büsing, C. and D'Andreagiovanni, F.: *Robust optimization under Multi-band uncertainty – Part I: Theory*, 2013.

[8] Büsing, C. and D'Andreagiovanni, F.: A New Theoretical Framework for Robust Optimization under Multi-band Uncertainty. In: *Operations Research Proceedings*, 2014, 115-121.

[9] Dantzig, G.B.: Linear programming under uncertainty. *Management Science* **1** (1955), 197-206.

[10] Gabrel, V., Lacroix, M., Murat, C. and Remli, N.: Robust location transportation problems under uncertain demands. *Discrete and Applied Mathematics* **164**, (2014), 100-111.

[11] Hitchcock, F.L.: The Distribution of a product from several sources to numerous localities. *Journal of Mathematics and Physics* **20**, (1941), 224-230.

[12] Magionni, F., Potra, F. and Bertocchi, M.: Stochastic versus robust optimization for a transportation problem. *Optimization online*, (2015).

[13] Monge, G.: Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, (1781), 666–704.

[14] Palma, C.D. and Nelson, J.D.:A robust optimization approach protected harvest scheduling decisions against uncertainty. *Canadian Journal of Forest Research*, **39**, (2009), 342-355.

[15] Soyster, A.L.: Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, **21**, (1973), 1154-1157.

Combining Estimates of Industry Production with the Structure of Input-Output Table

Vladimír Holý[1]

Abstract. Annual or quarterly production of specific industry can be estimated using univariate time series model. To capture and utilize the dependence between industries, multivariate time series model can be applied. Our approach deals with the dependency using input-output table of production flow between industries. Firstly, we estimate univariate production for each industry. Then we find the optimal combination of univariate estimates and estimates based on input-output structure.
We illustrate our model on the quarterly industry production of the Czech Republic. We estimate quarterly industry-by-industry input-output tables from annual input-output table using RAS method. Quarterly time series of production are modeled as seasonal ARIMA. Finally, we apply our multivariate model on estimates of input-output tables and productions. We compare results of our model with univariate time series forecasts.

Keywords: inter-industry relationships, input-output analysis, combination of estimates.

JEL classification: C32, C51, C67
AMS classification: 91B84

1 Introduction

In national economy, the production of industries and relations between them can be captured by input-output tables [7, 8]. Historical input-output tables are often estimated by RAS method [6]. Another problem is to forecast future input-output tables as their values may change over time [3]. This can be done for example by classical time series methods [9].

In this paper, we focus on forecasting of total productions of industries (i.e. row and column sums of input-output tables). We start with description of used notation in Section 2. We present the model for production forecasting in Section 3 and the method for input-output table estimation in Section 4. In Section 5 we propose a method which combines both approaches. We apply the proposed method to quarterly input-output tables of the Czech Republic in Section 6. Section 7 concludes with brief discussion.

2 Input-Output Framework

Let us assume we have n industries in national economy. Production of each industry and relations between industries can be captured by input-output equations

$$\begin{aligned} p_{t,i} &= \sum_{j=1}^{n} x_{t,i,j} + x_{t,i,0}, \qquad t = 1,\dots,s, \quad i = 0,\dots,n, \\ p_{t,j} &= \sum_{i=1}^{n} x_{t,i,j} + x_{t,0,j}, \qquad t = 1,\dots,s, \quad j = 0,\dots,n \end{aligned} \tag{1}$$

where $x_{t,i,j}$ is the inter-industry sale from industry $i = 1,\dots,n$ to industry $j = 1,\dots,n$ in period t, $x_{t,i,0}$ is the final demand of industry $i = 1,\dots,n$ in period t, $x_{t,0,j}$ is the value added of industry $j = 1,\dots,n$ in period t and $p_{t,k}$ is the total production of industry k in period t. We set $x_{t,0,0} = 0$. The first equation in (1) describes output of industry i while the second equation describes input of industry j. Furthermore, we denote

$$\begin{aligned} a_{t,i,j} &:= \frac{x_{t,i,j}}{p_{t,j}}, \qquad t = 1,\dots,s, \quad i = 0,\dots,n, \quad j = 0,\dots,n, \\ b_{t,i,j} &:= \frac{x_{t,i,j}}{p_{t,i}}, \qquad t = 1,\dots,s, \quad i = 0,\dots,n, \quad j = 0,\dots,n. \end{aligned} \tag{2}$$

Coefficients $a_{t,i,j}$ are commonly reffered to as technical coefficients [7, 8]. The structure of input-output table in our notation is presented in Table 1.

[1]University of Economics, Prague, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czech Republic, vladimir.holy@vse.cz

Input-Output Table	Intermediate Consumption			Final Demand	Total Production
Intermediate Consumption	$x_{t,1,1}$	$\cdots$	$x_{t,1,n}$	$x_{t,1,0}$	$p_{t,1}$
	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
	$x_{t,n,1}$	$\cdots$	$x_{t,n,n}$	$x_{t,n,0}$	$p_{t,n}$
Value Added	$x_{t,0,1}$	$\cdots$	$x_{t,0,n}$	0	$p_{t,0}$
Total Production	$p_{t,1}$	$\cdots$	$p_{t,n}$	$p_{t,0}$	

Table 1 The structure of input-output table describing n industries in period t.

3 Production Time Series Estimation

We model total production $p_{t,k}$ as seasonal ARIMA$(p,d,q)(P,D,Q)_S$ model. We set $S = 4$ as we use quarterly time series. For each industry we choose appropriate parameters p, d, q and seasonal parameters P, D, Q. Differecing parameters d, D are selected according to Canova-Hansen test [2]. Autoregressive parameters p, P and moving average parameters q, Q are selected by minimizing Akaike information criterion. This approach follows the automatic forecasting algorithm of [5].

4 Input-Output Table Estimation

Input-output tables are often compiled with lower frequency than the total production. For example, in the Czech Republic input-output tables are compiled in a five year period while total productions are recorded quarterly. To estimate annual tables in missing years we use RAS method [6, 7]. It takes last known input-output table and iteratively adjusts its elements to be consistent with given row and column sums (i.e. total productions). Similar approach is used for estimation of quarterly tables. The last known elements of table for given quarter are adjusted to be consistent with given quarterly total productions. However, this does not ensure that quarterly tables add up to annual table. To deal with this problem we use multidimensional RAS method, which besides row and column sums also ensures sums to annual table [4].

For forecasting quaterly coefficients $a_{t,i,j}$ and $b_{t,i,j}$ we simply take the values of last quaterly table (i.e. the estimated quarterly table with known total productions). Although the coefficients change over time, in short horizon (in our case 1 year) can yield reasonable predictions [3].

5 Combination of Estimates

We now combine time series estimates of total production with input-output table estimates. Our goal is to find weighted combination of estimates with weights φ_k for time series estimates, α_k for estimates based on output structure and β_k for estimates based on input structure. The weights are selected to minimize least squares of combined estimates of total production and real values of total production in training sample. The optimization problem is given by

$$
\begin{aligned}
\underset{\alpha_k,\beta_k,\varphi_k}{\text{minimize}} \quad & \frac{1}{s}\sum_{t=1}^{s}\frac{1}{n+1}\sum_{k=0}^{n}\left(p_{t,k}^{\text{real}}-\hat{p}_{t,k}^{\text{comb}}\right)^2 \\
\text{such that} \quad & \hat{p}_{t,k}^{\text{comb}} = \alpha_k\sum_{j=0}^{n}\hat{a}_{t,k,j}\hat{p}_{t,j}^{\text{init}} + \beta_k\sum_{i=0}^{n}\hat{b}_{t,i,k}\hat{p}_{t,i}^{\text{init}} + \varphi_k\hat{p}_{t,k}^{\text{est}}, \qquad t=1,\ldots,s, \quad k=0,\ldots,n, \\
& \alpha_k+\beta_k+\varphi_k = 1, \qquad k=0,\ldots,n, \\
& \alpha_k \geq 0, \qquad k=0,\ldots,n, \\
& \beta_k \geq 0, \qquad k=0,\ldots,n, \\
& \varphi_k \geq 0, \qquad k=0,\ldots,n
\end{aligned}
\tag{3}
$$

where α_k, β_k, φ_k are unknown weights, $p_{t,k}^{\text{real}}$ are known real values of total production, $\hat{p}_{t,k}^{\text{est}}$ are known time series estimates of total produciton and $\hat{p}_{t,k}^{\text{init}}$ are known initial estimates of total production. Resulting combined estimates of total production are denoted as $\hat{p}_{t,k}^{\text{comb}}$. This is a linearly constrained quadratic optimization problem.

The optimization problem (3) is solved repeatedly for different values of $\hat{p}_{t,k}^{\text{init}}$. We start by setting $\hat{p}_{t,k}^{\text{init}[0]} = \hat{p}_{t,k}^{\text{est}}$ resulting in combined estimates $\hat{p}_{t,k}^{\text{comb}[0]}$. In each following iteration i, we set $\hat{p}_{t,k}^{\text{init}[i]} = \hat{p}_{t,k}^{\text{comb}[i-1]}$ resulting in combined estimates $\hat{p}_{t,k}^{\text{comb}[i]}$. We stop iterations when change in weights is smaller than given threshold value.

Model	MAE	RMSE	Av. α_k	Av. β_k	Av. φ_k
ARIMA	4.3782	17.7401	0.0000	0.0000	1.0000
COMB–1Y	3.7781	16.6483	0.2986	0.2783	0.4232
COMB–2Y	3.8702	16.9846	0.2980	0.3073	0.3947
COMB–5Y	4.1640	18.3507	0.2769	0.3494	0.3737

Table 2 Errors and average values of weights for 2010Q1–2015Q4 forecasts.

In our case this approach leads to computationaly easy problem as the convergence is achieved within first few iterations.

A more direct alternative would be to modify optimization problem (3) by replacing $\hat{p}^{\text{init}}_{t,k}$ with $\hat{p}^{\text{comb}}_{t,k}$. However, this would require $\hat{p}^{\text{comb}}_{t,k}$ to be variables in the model rather than just a notation. Constraints in this model would no longer be linear as they would include multiples of variables $\hat{p}^{\text{comb}}_{t,k}$ and α_k, β_k. The resulting problem would be quadratically constrained quadratic program, which is NP-hard problem [1]. Due to computational difficulty we do not use this approach in our empirical study and leave it for future research.

Future intermediate consumption, final demand or value added can be forecasted by applying RAS method (or multidimensional RAS method) to last known input-output structure and proposed forecasts of total production (row and column sums).

6 Empirical Results

In empirical study we analyze the quarterly production of the Czech Republic economy split into $n = 41$ industries. We have available quarterly data of total production from the year 1995 to the year 2015 and annual input-output tables for years 1995, 2000, 2005, 2010 and 2013. Data are provided by Czech statistical office.

For estimation of seasonal ARIMA parameters we use 10 year history. For quarterly input-output tables we use coefficients $a_{t,i,j}$, $b_{t,i,j}$ from the same quarter of previous year. We consider 1 year (COMB-1Y), 2 year (COMB-2Y) and 5 year (COMB-5Y) history for optimization problem (3) (i.e. for finding the optimal weights). We compare the results with real values on testing period from 2010Q1 to 2015Q4 based on mean absolute deviation

$$MAE := \sum_{t=1}^{s}\sum_{k=1}^{n}\left|p^{\text{real}}_{t,k} - \hat{p}^{\text{comb}}_{t,k}\right| \tag{4}$$

and root mean squared error

$$RMSE := \sqrt{\sum_{t=1}^{s}\sum_{k=1}^{n}\left(p^{\text{real}}_{t,k} - \hat{p}^{\text{comb}}_{t,k}\right)^2}. \tag{5}$$

For example, when forecasting 2015Q1 values with COMB-1Y model, we use time series of total production from 2004Q1 to 2013Q4 to estimate ARIMA parameters and technical coefficients from input-output table of 2013Q1. The year 2014 is then used for finding the optimal combination of estimates.

Table 2 shows that incorporating input-output structure can increase accuracy of forecasts. All three combination models have smaller MAE than the original ARIMA estimates. Both COMB-1Y and COMB-2Y also have smaller RMSE. Model COMB-5Y have higher RMSE due to inaccurate estimates in years 2013 and 2014 as we can see in Figures 1 and 2. Interestingly, the shorter history for finding optimal weights yields better results. This could be because of changes in preferences between time series estimates and input-output estimates over time. This hypothesis is also supported by Table 3, in which we report mean absolute deviations of weights trained in different years. For 1 year apart the MAE can go up to 0.30 while for 5 year apart it can go up to 0.45 suggesting weights to be rather unstable in time. Figure 3 shows that majority of estimated weights are either very close to 0 or to 1. This suggest that proposed method selects the best estimation method for each industry rather than combining all estimation methods.

7 Discussion

We propose a method for forecasting of total productions which is a combination of time series estimates and estimates based on input-output structure. We apply this forecasting method on quarterly total production of 41 industries of the Czech Republic economy. Results show that proposed method have higher accuracy in the sense of MAE a RMSE over regular time series forecasts.

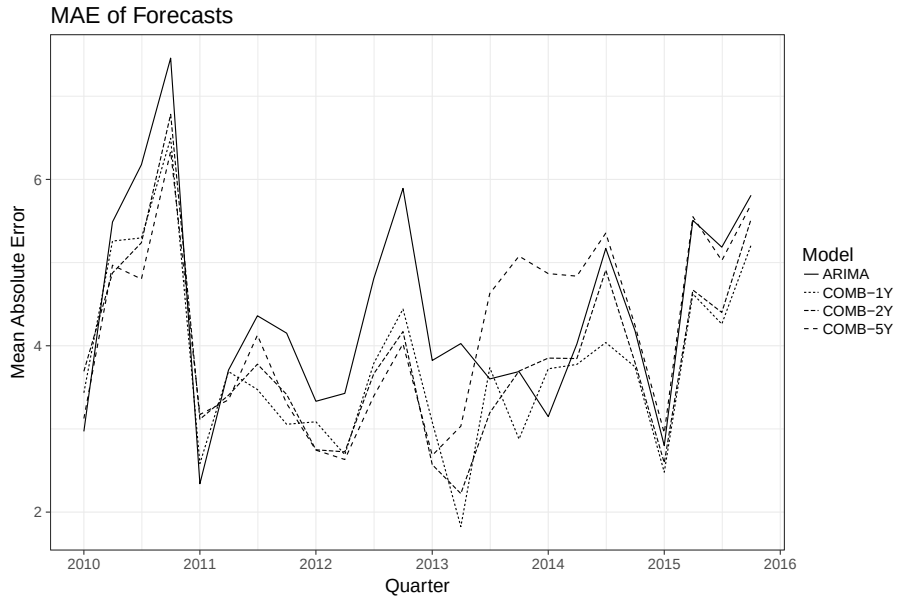


Figure 1 Mean absolute errors of production for 2010Q1–2015Q4 forecasts.

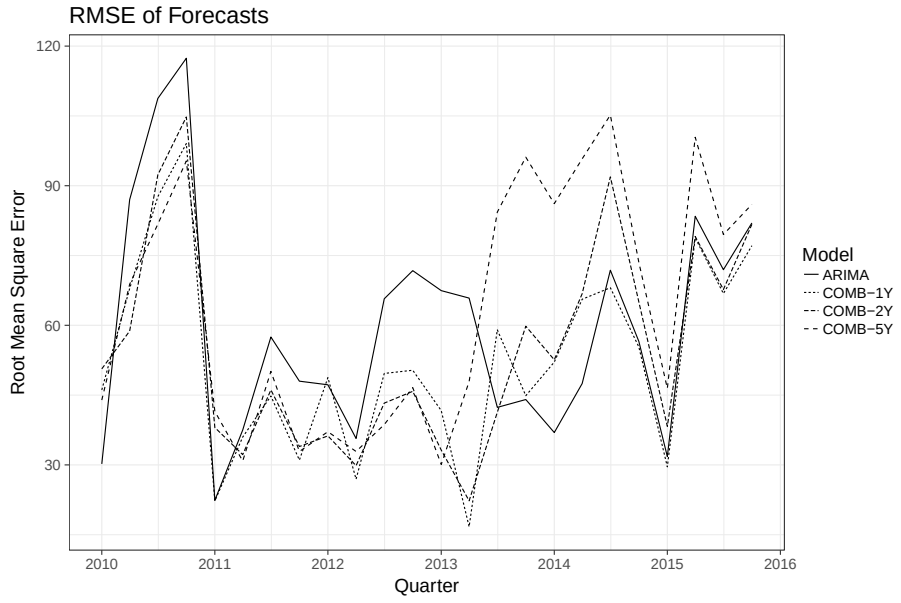


Figure 2 Root mean square errors of production for 2010Q1–2015Q4 forecasts.

	2010Q1	2011Q1	2012Q1	2013Q1	2015Q1	2015Q1
2010Q1	0.0000	0.2757	0.3107	0.3545	0.4106	0.4491
2011Q1	0.2757	0.0000	0.2686	0.3184	0.3665	0.3899
2012Q1	0.3107	0.2686	0.0000	0.3022	0.3348	0.3824
2013Q1	0.3545	0.3184	0.3022	0.0000	0.2932	0.3428
2014Q1	0.4106	0.3665	0.3348	0.2932	0.0000	0.2944
2015Q1	0.4491	0.3899	0.3824	0.3428	0.2944	0.0000

Table 3 Mean absolute errors of weights for first quarters of COMB–Y1 model.

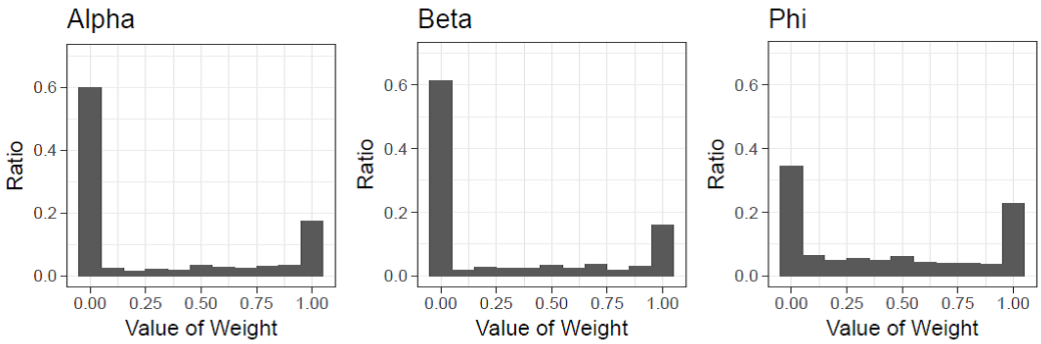


Figure 1 Histograms of estimated weights α_k, β_k and ϕ_k.

We identify two problems for future research. Firstly, the optimization problem for finding weights can be modified to capture relations between estimates more directly. However, this would require the problem of finding optimal weights to be quadratically constrained quadratic program. Secondly, results show that estimated optimal weights are not very stable in time. Finding suboptimal but more stable weights could further increase forecasting accuracy. It is also possible that these two problems are connected.

Acknowledgements

The work on this paper was supported by the grant IGS F4/28/2017 of University of Economics, Prague.

References

[1] Albers, C. J., Critchley, F., and Gower, J. C.: Quadratic Minimisation Problems in Statistics. *Journal of Multivariate Analysis* **102** (2011), 698-713. Retrieved from https://doi.org/10.1016/j.jmva.2009.12.018.

[2] Canova, F., and Hansen, B. E.: Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability. *Journal of Business and Economic Statistics* **13** (1995), 237-252. Retrieved from https://doi.org/10.2307/1392184.

[3] Haldi, J.: A Test on Two Hypotheses Concerning Interpretation of Input-Output Coefficients. *Visual Geosciences* **83** (1959), 1-14. Retrieved from https://doi.org/10.2307/40434578.

[4] Holý, V., and Šafr, K.: The Use of Multidimensional RAS Method in Input-Output Matrix Estimation, 2017. Working paper. Retrieved from https://arxiv.org/abs/1704.07814.

[5] Hyndman, R. J., and Khandakar, Y.: Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software* **27** (2008), 1-22. Retrieved from https://doi.org/10.18637/jss.v069.i12.

[6] Lahr, M. L., and de Mesnard, L.: Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis. *Economic Systems Research* **16** (2004), 115-134. Retrieved from https://doi.org/10.1080/0953531042000219259.

[7] Miller, R., and Blair, P.: *Input-Output Analysis: Foundations and Extensions.* 2nd edition. Cambridge University Press, New York, 2009. Retrieved from https://doi.org/10.1017/cbo9780511626982.

[8] ten Raa, T.: *The Economics of Input-Output Analysis*. Cambridge University Press, New York, 2006. Retrieved from https://doi.org/10.1017/cbo9780511610783.

[9] Sawyer, J. A.: Forecasting with Input-Output Matrices: Are the Coefficients Stationary? *Economic Systems Research* **4** (1992), 325-348. Retrieved from https://doi.org/10.1080/09535319200000030.

Critical period method for approximate solution of a discrete discounted stochastic program

Milan Horniaček[1]

Abstract. We develop an algorithm for approximate solution of an infinite horizon discrete time stochastic program with discounting of future single period payoffs. Uncertainty is captured by a Markov chain with countable infinite state space. In each period only a finite number of states occur with a positive probability. Finite sets of single period feasible actions and single period payoffs from them depend only on the current state. A strategy assigns to each state a feasible action at it. An epsilon-optimal strategy gives the sum of discounted expected single period payoffs that is no lower than its maximum on the set of strategies minus epsilon. Since the state space is infinite, an epsilon-optimal solution to the program is open-ended. We develop an algorithm that generates (for a chosen small positive epsilon) an epsilon-optimal solution gradually. It gives for each state the prescription for it that is ex ante epsilon-optimal, without specifying prescriptions for other states. It is based on critical periods. A critical period of a current state is the first period with the property that expected payoff in it cannot affect epsilon-optimality of action at the current state.

Keywords: discounted stochastic program, epsilon-solution, critical period method.

JEL Classification: C61
AMS Classification: 90C40

1 Introduction

Stochastic programs are widely used in decision-making. If they include development of random factors affecting outcomes of decisions, they usually have the form of a Markov decision process. (See, for example, [1], subchapter 1.5 for their characterization and Chapter 5 for examples of applications.) If there is not a final date of existence of an object of decision-making, they have infinite horizon, usually with discounting of future single period payoffs. Even in this case the set of states can be finite. (See [2] for an example.) Nevertheless, if decisions in future periods modify still prevailing effects of previous decisions, then states express cumulative effects of several decisions and the set of states is infinite. This includes, for example, programs with decisions on advertising, investment, or (product or technological) innovation by a monopolist or a perfectly competitive firm, or models with use of durable goods. In this case solution to a stochastic program is open-ended. The whole solution can never be computed and recorded. For each state we have to compute the prescription of an optimal strategy for it, without computing prescriptions for all following states.

Model predictive control ([3], [4]) takes this approach. It replaces an infinite horizon stochastic program (or a stochastic program with a long finite time horizon) with a sequence of finite horizon programs (or programs with a shorter time horizon). Nevertheless, solutions obtained by it need not be ex ante optimal or at least ε-optimal (i.e. optimal or ε-optimal with respect to available information about random factors affecting the system in the future). Moreover, there can be some technical problems with its application. In its basic form, it assumes that finite horizon problems are time invariant ([3], p. 790). This assumption is violated, for example, if a firm minimizes some sort of costs or maximizes profit and prices or some parameters of inverse demand curve (that it cannot effect) change over time. Schildbach, and Morari [4] do not make such assumption but they assume that the resulting optimization problem would become convex if all uncertain variables were known and fixed ([4], Assumption 1, p. 543). This assumption is not satisfied if sets of feasible decisions are finite.

We take a different approach. We restrict attention to ε-optimal strategies. For $\varepsilon > 0$, an ε-optimal strategy gives the sum of discounted expected single period payoffs that is no lower than its maximum on the set of strategies minus ε. For each state we compute the prescription of an ε-optimal strategy (for the whole infinite horizon program) for it, without computing prescriptions for following states. This is possible because (besides discounting of future single period payoffs) we assume finite sets of feasible actions and actions taken in the current period affect expected values of state variables in all following periods. The assumption that sets of feasible actions are finite can be justified by three observations. First, there is usually some smallest measurement unit that makes sense in a modeled environment (e.g. it does not make sense to consider smaller monetary units

[1] Comenius University in Bratislava, Faculty of Social and Economic Sciences, Institute of Economics, Mlynské luhy 4, SK-82105 Bratislava, milan.horniacek@fses.uniba.sk

than cents, or in production of beverages there is no need to consider smaller volume units than milliliters). Second, numerical representations of decisions are naturally bounded (e.g. by the fact that they cannot be negative and expenditures on a decision cannot exceed some pre-specified budget). Third, decisions are made with the help of digital computers. These machines can represent only a finite number of digits in both integer and decimal part of a number.

For each state we compute the prescription of an ε-optimal strategy for it by solving a finite succession of linear programs, in which we take into account only first *n* periods and states that can occur with a positive probability (under some strategy) in them. At each state that can be reached from the states taken into account we specify the highest feasible value of each action and use infinite repetition of the payoff gross of cost of taking actions at state generated by them to replace the value of strategy at it (i.e. the sum of discounted expected single period payoffs starting from it). We end this successive solving of linear programs in period *n*, which is critical, i.e. it is the first period with the property that expected payoff in it cannot affect epsilon-optimality of action at the current state.

Throughout the paper, N denotes the set of positive integers and $\Re$ is the set of real numbers. We endow each finite dimensional real vector space with the Euclidean topology and each infinite dimensional product of finite dimensional spaces with the product topology.

2 Stochastic program

Time horizon of the analyzed program is N. Ω is the countable infinite set of states. Each $\omega \in \Omega$ is an element of the Euclidean space with dimension m^*. For each $\omega \in \Omega$, $X(\omega) = \prod_{i=1}^{m^*} X_i(\omega)$ is the finite non-singleton set of feasible actions at ω. The symbol $\rho(\omega, x)$ stands for the probability distribution on Ω in the following period when state in the current period is ω and action x is taken at it; $\rho(\omega, x)(\omega')$ is the probability of occurrence of state ω' in the following period. We assume that the support of $\rho(\omega, x)$ is finite and it has at least two elements.

Assumption 1. *For each $i \in \{1,\ldots,m^*\}$ and each $\tau \in N$, the expected value of ω_i reached τ periods after the current period strictly increases in action x_i taken in the current period and does not decrease in action x_j, $j \in \{1,\ldots,m^*\} \setminus \{i\}$, taken in the current period.*

Function $\pi : \{(\omega, x) \mid \omega \in \Omega, x \in X(\omega)\} \to L$, where L is an infinite compact subset of $\Re$, assigns a single period payoff to each state ω and action x taken at it. Future single period payoffs are discounted (without discounting the current period payoff) using discount factor $\delta \in (0,1)$.

Assumption 2. *For each (ω, x) from domain of π, (a) $\pi(\omega, x) = \eta(\omega) - \gamma(\omega, x)$, where $\eta(\omega)$ is a part of payoff at state ω that is independent of x and $\gamma(\omega, x)$ is the cost of taking an action x at state ω, (b) $\eta(\omega)$ is strictly increasing in ω_i for each $i \in \{1,\ldots,m^*\}$, and (c) $\gamma(\omega, x)$ is strictly increasing in x_i for each $i \in \{1,\ldots,m^*\}$, (d) for each $\omega \in \Omega$, $0 \in X(\omega)$ (where $0 \in \Re^{m^*}$) and $\gamma(\omega, 0) = 0$.*

We restrict attention to Markov strategies. In the analyzed stochastic program only current state is payoff relevant. (The set of feasible actions in the current period and their payoff consequences depend only on the current state.) Therefore, a Markov strategy (henceforth, "strategy") is a function that assigns to each $\omega \in \Omega$ an element of $X(\omega)$. We use the symbol Q for the set of strategies and q for its generic element.

We denote by $\Omega^{(k)}(\omega')$ the set of states that can occur (under some strategy) with a positive probability in period $k \in N$ when the initial state is ω'. Since sets of feasible actions are finite and $\rho(\omega, x)$ has a finite support, $\Omega^{(k)}(\omega')$ is finite. It is defined inductively by

$$\Omega^{(1)}(\omega') = \{\omega'\} \tag{1}$$

and

$$\Omega^{(k)}(\omega') = \{\omega \in \Omega \mid \exists \omega'' \in \Omega^{(k-1)}(\omega') \,\&\, x \in X(\omega'') \text{ such that } \rho(\omega'', x)(\omega) > 0\},\ k \in N \setminus \{1\}. \tag{2}$$

For $n \in N$ let $\bar{\Omega}^{(n)}(\omega')$ be the set of states that can occur (under some strategy) with a positive probability in the first n periods when the initial state is ω'. Thus,

$$\bar{\Omega}^{(n)}(\omega') = \bigcup_{k=1}^{n} \Omega^{(k)}(\omega'). \tag{3}$$

Function $\mu : \Omega \times Q \times N \times \Omega \to [0,1]$ assigns to each (ω', q, k, ω) the probability $\mu(\omega', q, k, \omega)$ that state ω occurs under strategy q in period k when the initial state is ω'. It is defined inductively by

$$\mu(\omega', q, 1, \omega') = 1, \quad \mu(\omega', q, 1, \omega) = 0 \quad \forall \omega \in \Omega \setminus \{\omega'\} \tag{4}$$

and

$$\mu(\omega', q, k, \omega) = \sum_{\omega'' \in \Omega^{(k-1)}(\omega')} \mu(\omega', q, k-1, \omega'') \rho(\omega'', q(\omega))(\omega), \qquad \forall k \in N \setminus \{1\}. \tag{5}$$

The analyzed stochastic program with initial state $\omega' \in \Omega$ maximizes the sum of discounted expected single period payoffs over the infinite horizon. It has the form

$$\max_{q \in Q} u(q) = \sum_{t \in N} \left(\delta^{t-1} \sum_{\omega \in \Omega^{(t)}(\omega')} \mu(\omega', q, t, \omega) \pi(\omega, q(\omega)) \right). \tag{6}$$

Since future single period payoffs are discounted and the range of function π is bounded, the objective function in (6) is well-defined. It can be shown that it is continuous.

Q is a countable infinite product of finite (and, hence, compact) subsets of Euclidean spaces, so it is compact by Tychonoff's theorem. Thus, program (6) is a maximization of a continuous function on a non-empty compact set, so it has an optimal solution.

Let u^* be the global maximal value of the objective function in (6) and $\varepsilon > 0$ be a small number.

Definitin 1. *A strategy $q \in Q$ is an ε-optimal solution of stochastic program (6) if $u^* - u(q) < \varepsilon$.*

3 Algorithm

For $\omega \in \Omega$ let v_ω be the value of a strategy $q^* \in Q$ at state ω (i.e. the sum of discounted expected single period payoffs from a strategy q^* when the initial state is ω. That is,

$$v_\omega = \pi(\omega, q^*(\omega)) + \delta \sum_{\omega' \in \Omega^{(2)}(\omega)} \rho(\omega, q^*(\omega))(\omega') v_{\omega'}. \tag{7}$$

Applying Bellman's principle of optimality, a strategy $q^* \in Q$ is an optimal solution of program (6) if and only if for each $\omega \in \Omega$

$$v_\omega = \max_{x \in X(\omega)} \left(\pi(\omega, x) + \delta \sum_{\omega'' \in \Omega^{(2)}(\omega)} \rho(\omega, x)(\omega'') v_{\omega'} \right). \tag{8}$$

Take (arbitrary) $\omega' \in \Omega$. In order to compute prescriptions of ε-optimal solutions of program (6) at state ω', we proceed in a finite number of iterations. For $n \in N$, let $X^{(n)}(\omega')$ be a nonempty finite subset of $X(\omega')$ - a subset of $X(\omega')$ containing feasible actions at ω' that are considered in iteration n, and $Y^{(n)}(\omega') \subseteq X(\omega')$ be a set of prescriptions of ε-optimal solutions of program (6) at state ω' computed in preceding iterations. We set $X^{(1)}(\omega') = X(\omega')$ and $Y^{(1)}(\omega') = \varnothing$. (Sets $X^{(n)}(\omega')$ and $Y^{(n)}(\omega')$ for $n > 1$ are defined below.)

In iteration $n \in N$ we consider only states in $\bar{\Omega}^{(n)}$ and set

$$v_\omega = \eta(\omega) + \frac{\delta}{1-\delta} \sum_{\omega'' \in \Omega^{(n+2)}(\omega')} \rho\left(\omega, \left(\max X_i(\omega)\right)_{i \in \{1, \ldots, m^*\}}\right)(\omega'') \eta(\omega''), \quad \forall \omega \in \Omega^{(n+1)}(\omega') \setminus \bar{\Omega}^{(n)}(\omega'). \tag{9}$$

That is, for each state ω that can occur with a positive probability in period $n+1$ but not in the first n periods, we replace the value of a strategy at ω by the sum of two values. The first one equals single period payoff at ω gross off cost of an action taken at it. The second one equals the sum of discounted single period payoffs gross off cost of taking actions starting from the following period that would result if the expected single

period payoff in the following period generated by taking the highest feasible value of each action at ω was infinitely repeated. (This is not the only way how to define v_ω for $\omega \in \Omega^{(n+1)}(\omega') \setminus \bar{\Omega}^{(n)}(\omega')$. We use it because it is relatively simple and it would be useful if we wanted to compute optimal strategies rather than ε-optimal ones.)

For each $x' \in X^{(n)}(\omega')$ we replace (8) by

$$v_{\omega'} = \pi(\omega', x') + \delta \sum_{\omega'' \in \Omega^{(2)}(\omega')} \rho(\omega', x')(\omega'') v_{\omega''}, \tag{10}$$

$$v_\omega = \max_{x \in X(\omega)} \left(\pi(\omega, x) + \delta \sum_{\omega'' \in \Omega^2(\omega)} \rho(\omega, x)(\omega'') v_{\omega''} \right), \qquad \forall \omega \in \overline{\Omega}^{(n)} \setminus \{\omega'\},$$

and (9). Thus, we want to compute a strategy $q^{(n)}|x'$ that assigns actions to states in $\bar{\Omega}^{(n)}$ and maximizes the sum of discounted expected single period payoffs over infinite horizon subject to two constraints. First, it assigns action x' to state ω'. Second, its value at each state $\omega \in \Omega^{(n+1)}(\omega') \setminus \bar{\Omega}^{(n)}(\omega')$ is given by (9).

Such strategy can be computed from a solution of the linear program

$$\min \sum_{\omega \in \bar{\Omega}^{(n)}(\omega')} v_\omega \tag{11}$$

subject to (9)-(10) and

$$v_\omega \geq \pi(\omega, x) + \delta \sum_{\omega'' \in \Omega^{(2)}(\omega)} \rho(\omega, x)(\omega'') v_{\omega''}, \qquad \forall \omega \in \bar{\Omega}^{(n)}(\omega') \setminus \{\omega'\}, \qquad \forall x \in X(\omega). \tag{12}$$

Let $v_{\omega'}^{(n)}(x)$ be the value of $v_{\omega'}$ in an optimal solution of linear program (9)-(12) and set

$$v_{\omega'}^{(n)*} = \max_{x \in X^{(n)}(\omega')} v_{\omega'}^{(n)}(x).$$

That is, $v_{\omega'}^{(n)*}$ is the maximal sum of discounted expected single period payoffs over the infinite horizon, when the initial state is ω', subject to the use of (9) for each state $\omega \in \Omega^{(n+1)}(\omega') \setminus \bar{\Omega}^{(n)}(\omega')$.

If $x^* \in X^{(n)}(\omega')$ is the prescription of each optimal solution of linear program (9)-(12) at state ω', we set

$$v_{\omega'}^{(n)+} = \max\left\{ v_{\omega'}^{(n)}(x) \,\middle|\, x \in X^{(n)}(\omega'), v_{\omega'}^{(n)}(x) < v_{\cdot}^{(n)*} \right\}. \tag{13}$$

If, in addition,

$$v_{\omega'}^{(n)*} - v_{\omega'}^{(n)+} > \frac{2\delta^n \max L}{1-\delta} - \varepsilon, \tag{14}$$

then x^* is the prescription of some ε-optimal strategy of the stochastic program (6) with initial state ω' at state ω'. (If (14) holds also for $\varepsilon = 0$, then x^* is the prescription of each optimal strategy of the stochastic program (6) with initial state ω' at state ω'.) This can be seen as follows. In (14) we decreased the sum of discounted expected single period payoffs starting from period $n+1$ from a strategy solving program (9)-(12) by $\frac{\delta^n \max L}{1-\delta}$, which is an upper bound on such decrease (recall part (d) of Assumption 2), and increased the sum of discounted expected single period payoffs starting from period $n+1$ from a strategy giving payoff over the infinite horizon equal to $v_{\omega'}^{(n)+}$ by $\frac{\delta^n \max L}{1-\delta}$, which is an unattainable upper bound on such increase (recall that cost of taking actions are disregarded and highest feasible values of actions are used in (9)). Of course, moving from program (9)-(12) to program (6), the sum of discounted expected single period payoffs in the first n periods from both approaches can change. Nevertheless, an increase in the latter sum would involve lower cost of actions and, hence, lower values of actions. (If it was achieved only by a change in distributions of probabilities of states in the following periods, it would be attainable also in a solution of (9)-(12).) This would decrease the sum of discounted expected single period payoffs since period $n+1$. Since the changes in this sum for both approaches used in (14) are on their upper bounds, any increase in the sum of discounted expected single period

payoffs in the first n periods from a strategy giving payoff over the infinite horizon equal to $v_{\omega'}^{(n)+}$ is compensated in an increase in the sum of discounted expected single period payoffs considered in (14).

By an analogous reasoning, if for some $x \in X^{(n)}(\omega')$

$$v_{\omega'}^{(n)}(x) - v_{\omega'}^{(n)*} > \frac{2\delta^n \max L}{1-\delta} - \varepsilon, \tag{15}$$

then x is the prescription of some ε-optimal strategy of (6) with initial state ω' at state ω'. (In (15), we decreased the sum of discounted expected single period payoffs since period $n+1$ from a strategy prescribing x at ω' by $\frac{\delta^n \max L}{1-\delta}$ and increased by this amount the sum of discounted expected single period payoffs since period $n+1$ from a strategy giving payoff over the infinite horizon equal to $v_{\omega'}^{(n)*}$.) Note that (15) can hold only if (14) holds. Thus, if $x^* \in X^{(n)}(\omega')$ is the prescription of each optimal solution of linear program (9)-(12) at state ω' and (14) holds, we set

$$Y^{(n+1)}(\omega') = Y^{(n)}(\omega') \cup \{x^*\} \cup \{x \in X^{(n)}(\omega') \mid (15)\ holds\}. \tag{16}$$

If $x^* \in X^{(n)}(\omega')$ is the prescription of each optimal solution of linear program (9)-(12) but (14) does not hold, then $Y^{(n+1)}(\omega') = Y^{(n)}(\omega')$. If linear program has at least two solutions with different prescriptions at state ω', we set

$$Y^{(n+1)}(\omega') = Y^{(n)}(\omega') \cup \{x \in X^{(n)}(\omega') \mid (15)\ holds\}\}. \tag{17}$$

If for some $x \in X^{(n)}(\omega')$

$$v_{\omega'}^{(n)}(x) - v_{\omega'}^{(n)*} \le -\frac{2\delta^n \max L}{1-\delta} - \varepsilon, \tag{18}$$

then x cannot be the prescription of any ε-optimal strategy of (6) with initial state ω' at state ω'. In (18), we increased the sum of discounted expected single period payoffs since period $n+1$ from a strategy prescribing x at ω' by $\frac{\delta^n \max L}{1-\delta}$ and decreased by this amount the sum of discounted expected single period payoffs since period $n+1$ from a strategy giving payoff over the infinite horizon equal to $v_{\omega'}^{(n)*}$. If this is not enough to make (15) hold, then x cannot be the prescription of any ε-optimal strategy of (6) with initial state ω' at state ω'. Therefore, we set

$$X^{(n+1)}(\omega') = X^{(n)}(\omega') \setminus \{x \in X^{(n)}(\omega') \mid (18)\ holds\}. \tag{19}$$

When we reach $n \in N$ with $X^{(n)}(\omega') = Y^{(n)}(\omega')$, then $Y^{(n)}(\omega')$ is the set of all prescriptions of ε-optimal strategies of (6) with initial state ω' at state ω'. In this case period n is the critical period for program (6) with initial state ω'. We proceed with computations for some initial state $\omega'' \in \Omega$ such that, for some x that is the prescription of an ε-optimal strategy of (6) with initial state ω' at state ω', $\rho(\omega', x), (\omega'') > 0$ (unless we have already made computations for it). If the decision-maker can wait for advice until the actual state in the following period becomes known, we make computations only for this actual state. If $X^{(n)}(\omega') \supset Y^{(n)}(\omega')$, we proceed with iteration n.

Remark 1. It is true that we could use only (15) to define (17), without considering (13) and (14). Nevertheless, if in each iteration each solution of linear program (9)-(12) prescribes the same action x^* at state ω', we would have to wait until iteration n with $\frac{2\delta^n \max L}{1-\delta} < \varepsilon$ in order to identify x^* as the prescription of some ε-optimal strategy of (6) with initial state ω' at state ω'. The approach using (13) and (14) allows us to do it earlier.

Remark 2. It may seem that it would be enough to include only one element of $Y^{(n)}(\omega')$ in $X^{(n)}(\omega')$. Nevertheless, we do not know which element of $Y^{(n)}(\omega')$ gives $v_{\omega'}^{(n)*}$ (if any of them gives it). Since $v_{\omega'}^{(n)*}$ is used in (15), we need to have $Y^{(n)}(\omega') \subseteq X^{(n)}(\omega')$.

There is a finite $n \in N$ for which $Y^{(n)}(\omega') = X^{(n)}(\omega')$. This can be seen as follows. If x^* is the prescription of each optimal solution of (6) with initial state ω' at ω', then there is $n_1 \in N$ such that $v_{\omega'}^{(n)}(x^*) - v_{\omega'}^{(n)}(x) > -0.5\varepsilon$ for each $x \in X(\omega') \setminus \{x^*\}$ and each $n \geq n_1$. There is also $n_2 \in N$ such that $\frac{2\delta^n \max L}{1-\delta} \leq 0.5\varepsilon$ for each $n \geq n_2$. Then for each $n \geq max\{n_1, n_2\}$ (14) holds. Next suppose that q^* is an optimal strategy of (6) with initial state ω' prescribing action x^* at ω' and q is a strategy in (6) with initial state ω' prescribing action $x \neq x^*$ at ω' and satisfying $u(q) - u(q^*) = -\Delta \leq 0$. If $\Delta < \varepsilon$, then there exist $n_3 \in N$ and $n_4 \in N$ such that $v_{\omega'}^{(n)}(x) - v_{\omega'}^{(n)}(x^*) > -0.5(\Delta + \varepsilon)$ for each $n \geq n_3$ and $\frac{2\delta^n \max L}{1-\delta} \leq 0.5(\varepsilon - \Delta)$ for each $n \geq n_4$. Therefore, for each $n \geq max\{n_3, n_4\}$ (14) holds. If $\Delta \geq \varepsilon$, then for each $n \in N$ moving from $u(q)$ and $u(q^*)$ to $v_{\omega'}^{(n)}(x)$ and $v_{\omega'}^{(n)}(x^*)$ increases $v_{\omega'}^{(n)}(x) - v_{\omega'}^{(n)}(x^*)$ (if it increases it at all) by less than $\frac{2\delta^n \max L}{1-\delta}$, so that (18) holds.

The preceding considerations give rise to an algorithm given below. In its description we use the assignment sign ($\rightarrow$) whenever the use of equality sign ($=$) would be mathematically incorrect. Ω^+ is the set of states for which computations have already been made. We apologize to referees and readers that, due to space limitations, we cannot give here an example of application of the algorithm.

Algorithm.

Step 1. Set ω' equal to the initial state and $\Omega^+ = \varnothing$. Go to step 2.

Step 2. Set $n = 1$, $X^{(1)}(\omega') = X(\omega')$, and $Y^{(n)}(\omega') = \varnothing$. Go to step 3.

Step 3. Compute $v_{\omega'}^{(n)}(x)$ for each $x \in X^{(n)}(\omega')$. If $x^* \in X^{(n)}(\omega')$ is the prescription of each optimal solution of linear program (9)-(12), go to step 4. Otherwise, go to step 6.

Step 4. Compute $v_{\omega'}^{(n)+}$ according to (13). If (14) holds, compute $Y^{(n+1)}(\omega')$ according to (16) and go to step 7. Otherwise, go to step 5.

Step 5. Set $Y^{(n+1)}(\omega') = Y^{(n)}(\omega')$ and go to step 7.

Step 6. Compute $Y^{(n+1)}(\omega')$ according to (17) and go to step 7.

Step 7. Compute $X^{(n+1)}(\omega')$ according to (19) and go to step 8.

Step 8. If $Y^{(n+1)}(\omega') = X^{(n+1)}(\omega)$, set $\Omega^+ \cup \{\omega'\} \rightarrow \Omega^+$ and go to step 9. Otherwise, set $n+1 \rightarrow n$ and go to step 3.

Step 9. Choose $\omega'' \in \Omega \setminus \Omega^+$ such that $\rho(\omega', x)(\omega'') > 0$ for some $x \in Y^{(n+1)}(\omega')$. Set $\omega'' \rightarrow \omega'$ and go to step 2.

For the complexity of the algorithm the size of linear program (9)-(12) and number of iterations are crucial. The maximal number of constraints, as well as the maximal number of variables, in (9)-(12) depends linearly on the maximal number of states that can occur with a positive probability in any period. The maximal number of constraints depends linearly on the maximal number of actions that can be taken at any state. The number of iterations changes in a discontinuous way with respect to the discount factor. For each $\delta \in (0,1)$ a small increase in δ either leaves the number of iterations unchanged or increases it by one.

References

[1] Ching, W.K., Huang, X., Ng, M.K., Siu, T.K.: *Markov chains.* Springer, Heidelberg, 2013.

[2] Ching, W.K., Ng, M.K., Wong, K., Atlman, E.: Customer lifetime value: a stochastic programming approach. *Journal of Operations Research Society* **55** (2004), 860-868.

[3] Mayne, D.Q., Rawlings, J.B., Rao, C.V., Scokaert, P.O.M.: Constrained model predictive control: Stability and optimality. *Automatica* **36 (**2000), 789-814.

[4] Schildbach, G., Morari, M.: Scenario-based model predictive control for multi-echelon supply chain management. *European Journal of Operational Research* **252** (2016), 540-549.

The valuation of discretely sampled European lookback options: a DG approach

Jiří Hozman[1], Tomáš Tichý[2]

Abstract. Path-dependent options represent an important part of the derivatives traded in financial markets. One of the commonly used and popular subclasses of path-dependent options is formed by so-called lookback options with payoff depending on the minimum or maximum price of the underlying asset attained during the lifetime of the option and enabling the investors to sell at the highest or buy at the lowest price, i.e., the most favourable one. Commonly, the maximum or minimum are monitored at discrete dates so that there is no analytical pricing formulae and one has to rely on numerical techniques. In this paper we present a PDE approach to European lookback options leading to the usual Black-Scholes equation, where the path-dependent variable appears as a parameter only and discrete sampling is balanced by introducing the jump conditions across the sampling dates. Since the pricing equation is the same as for the plain vanilla option, the discontinuous Galerkin (DG) method is applied to the problem in the same manner, except for the treatment of jump conditions at each monitoring date. Finally, reference numerical experiments illustrate empirical findings.

Keywords: Option pricing, discontinuous Galerkin method, Black-Scholes model, path-dependent variable, lookback options, discrete sampling, jump conditions.

JEL classification: C44, G13
AMS classification: 35, 90C15

1 Introduction

A key condition of functioning financial markets is the knowledge of fair price of any traded security. In some cases (e.g., stocks or bonds) such price is given simply by matching the supply and demand, in other cases (e.g., options) it can be derived by no-arbitrage arguments. Hence, the aforementioned condition should be extended by the knowledges of efficient techniques for price calculation. It is thus not surprising that development of new techniques as well as deep analysis of currently known approaches is in the center of interest of many researchers since the seminal paper [2] on modern option pricing.

For example, as concerns lookback options, a special type of financial derivatives from the group of path-dependent options, payoff of which depends on the maximum or minimum asset price realized over some prespecified period (continuously or discretely monitored) of the option life, we can name [12], who focused on finite difference method, and [5], who focused on finite element methods; see also [1], [11] and the recent wavelet approach from [4] following paper [3].

In this short contribution we restrict ourselves to discretely monitored floating strike lookback put option and develop numerical scheme based on discontinuous Galerkin approach. This method is based on piecewise polynomial but discontinuous approximations, thus it allows us to solve degenerated parabolic PDE problems elegantly, among which we include the wide spectrum of option pricing equations. We proceed as follows – in Section 2 lookback options are classified, while in Section 3 we derive a relevant valuation PDE, including its conditions. In Section 4 a numerical procedure is constructed and finally, in Section 5 a numerical experiment is provided.

2 Classification of European lookback options

In this paper we consider only European-style lookback options, i.e., options exercising of which is permitted only at maturity time T. Such options exist either as put or call options and their value can be represented as a function of the actual time t, the underlying asset price $S(t)$ and one of the two path-dependent variables M and m denoting the maximum and minimum value of the realized asset price, respectively. These extremal values can be measured continuously on the whole time interval $[0,T]$, i.e., $M = \max_{0\le t\le T} S(t)$ and $m = \min_{0\le t\le T} S(t)$; or at the discrete times $\{t_i\}_{i=1}^N \subset [0,T]$ only, i.e., $M = \max(S(t_1), S(t_2), \ldots, S(t_N))$ and $m = \min(S(t_1), S(t_2), \ldots, S(t_N))$.

[1]Technical University of Liberec, Department of Mathematics and Didactics of Mathematics, Studentská 2, 461 17 Liberec, Czech Republic, jiri.hozman@tul.cz

[2]VŠB – Technical University of Ostrava, Department of Finance, Sokolská 33, 701 21 Ostrava, Czech Republic, tomas.tichy@vsb.cz

Then we speak of continuously or discretely sampled lookback options. Obviously, it always holds for continuous updating that the asset price is necessarily greater than or equal to the minimum and less or equal to the maximum. This is not true for discrete updating where the asset price may be out of these bounds.

Moreover, according to the way in which the variables M or m are incorporated into the payoff function, we distinguish four basic types of lookback options, see Table 1. A floating strike lookback put gives the holder the right to sell at the highest realized price while a floating strike lookback call allows him or her to buy at the lowest realized price over the lookback period. In other words, its payoff is similar to a vanilla option but with the corresponding path-dependent variable replacing the exercise price. On the other hand, a fixed strike lookback put (call) is a put (call) option on the minimum (maximum) realized price with agreed price K. Similarly, its payoff corresponds to a vanilla option with swapping the corresponding path-dependent variable for S.

payoff	put	call
floating strike	$\max(M - S(T), 0)$	$\max(S(T) - m, 0)$
fixed strike	$\max(K - m, 0)$	$\max(M - K, 0)$

Table 1 Payoff functions for four basic types of lookback options

The aforementioned classification identifies eight subtypes of lookback options. Due to the different treatment of particular subtypes, the rest of the paper is oriented on one of them only, for a complete overview see [11].

3 Discrete sampling of the maximum: the floating strike case

We concentrate on valuing a put option that depends on the maximum value M of the asset obtained by the discrete measurement, i.e., the maximum of finite set of real prices at which the underlying asset is traded. This type of sampling is more natural and describes the most lookback option contracts, especially for its easier calculation. Since the discretely measured maximum M does not change between sampling dates, the asset price can exceed it, i.e., $S > M$ is permissible. This is a very important difference between continuously and discretely sampled lookback options which is reflected on the (S, M)-domain on which the problem is posed.

In the case of discrete sampling the standard mathematical approach still leads to the Black-Scholes partial differential equation completed by the system of boundary and terminal (initial) conditions. Moreover, across the monitoring dates there is a jump condition guaranteeing the continuity of realized value of an option with respect to S, M and t.

3.1 Model dynamics, pricing equation and jump conditions

Let $V = V(S, M, t)$ denote the value of a floating strike lookback put option. Suppose that the price process $S(t)$ evolves over time according to this stochastic differential equation:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad S(0) > 0, \tag{1}$$

where $\mu S(t)dt$ is a drift term with a constant rate $\mu \geq 0$, $W(t)$ is a standard Brownian motion and $\sigma > 0$ is the volatility of the asset price. Unfortunately, the measurement of maximum M is not differentiable and thus has to be approximated by a path-dependent quantity:

$$M_n(t) = \left(\sum_{t_i < t} S(t_i)^n \right)^{1/n} = \left(\int_0^t \sum_{i=1}^{N} \delta(\tau - t_i) S(\tau)^n d\tau \right)^{1/n}, \tag{2}$$

where δ denotes the Dirac delta function. The derivative of this quantity is given by

$$dM_n(t) = \frac{1}{n} \sum_{i=1}^{N} \delta(t - t_i) \frac{S(t)^n}{M_n(t)^{n-1}} \, dt, \tag{3}$$

which implies that the dynamics of M_n is driven according to ordinary differential equation only, i.e., M_n is a deterministic variable. In the limit as $n \to \infty$ and for $t = T$ we obtain the maximum of the asset price taken at times $\{t_i\}_{i=1}^{N}$, i.e., $M = \lim_{n\to\infty} M_n = \max(S(t_1), S(t_2), \ldots, S(t_N))$.

Using the standard technique, (*i*) a construction of a hedged portfolio which contains one unit of a put option V payoff of which depends on M_n and $-\Delta$ units of the underlying asset S; (*ii*) an elimination of the stochastic components by setting $\Delta = \frac{\partial V}{\partial S}$; and (*iii*) a comparison of dynamics of the portfolio by virtue of multidimensional

Ito's lemma, one can demonstrate that the value of the option being priced must satisfy the following partial differential equation:

$$\frac{\partial V}{\partial t}+\frac{1}{n}\sum_{i=1}^{N}\delta(t-t_i)\frac{S^n}{(M_n)^{n-1}}\frac{\partial V}{\partial M_n}+\frac{1}{2}\sigma^2S^2\frac{\partial^2V}{\partial S^2}+rS\frac{\partial V}{\partial S}-rV=0,\ S>0,\ M_n>0,\ 0<t\le T, \quad (4)$$

where $r \geq 0$ is the risk-free interest rate. Since $\delta(t-t_i)=0$ for $t \notin \{t_i\}_{i=1}^N$, away from these times the equation (4) is insensitive to the presence of path-dependent variable M_n and it reduces to the classical Black-Scholes equation in S and t, with M (as $n \to \infty$) entering only as a parameter, i.e., S and M are considered as independent variables. Therefore, outside sampling dates, the pricing equation reads as

$$\frac{\partial V}{\partial t}+\frac{1}{2}\sigma^2S^2\frac{\partial^2V}{\partial S^2}+rS\frac{\partial V}{\partial S}-rV=0, \quad S>0,\ M>0,\ 0<t\le T. \quad (5)$$

The situation is completely different across sampling dates. Formally, the term proportional to the delta function grows beyond all bounds and can be balanced only by the time derivative of V. Near sampling dates the equation (4) is approximated by these dominant terms as the first order hyperbolic equation

$$\frac{\partial V}{\partial t}+\frac{1}{n}\sum_{i=1}^{N}\delta(t-t_i)\frac{S^n}{(M_n)^{n-1}}\frac{\partial V}{\partial M_n}=0, \quad S>0,\ M_n>0,\ t\in\{t_i\}_{i=1}^N. \quad (6)$$

The theory of the first order hyperbolic equations implies that the solution $V(S,M_n,t)$ of (6) is constant along the characteristics given by

$$\mathcal{H}(t-t_i)-\left(\frac{M_n}{S}\right)^n=constant, \quad (7)$$

where $\mathcal{H}$ is the Heaviside step function. Using this feature and following the approach from [11] we find that

$$\lim_{\varepsilon\to0+}V(S,M_n,t_i-\varepsilon)=\lim_{\varepsilon\to0+}V(S,S(1+(M_n/S)^n)^{1/n},t_i+\varepsilon), \quad i=1,\dots,N. \quad (8)$$

Then, in the limit case $n \to \infty$ the relation (8) goes to the jump condition across the sampling dates introduced via operator $\mathcal{J_C}$ as

$$\lim_{\varepsilon\to0+}V(S,M,t_i-\varepsilon)=\lim_{\varepsilon\to0+}\mathcal{J_C}\left(V(S,M,t_i+\varepsilon)\right):=\lim_{\varepsilon\to0+}V(S,\max(S,M),t_i+\varepsilon), \quad i=1,\dots,N, \quad (9)$$

where $\mathcal{J_C}$ is defined as the decision process in the limit transition of (8). It distinguishes between two cases, for $S \leq M_n$, the second argument of the right-hand side of (8) simply tends to M, however, for $S > M_n$, this argument tends to S. Note that (9) has also its financial reasoning that the realized value of the option must be continuous as S, M and t vary from arbitrage considerations.

3.2 Initial-boundary value problem

Since the pricing equation (5) with conditions (9) is accompanied by the payoff prescribed at maturity, the studied problem is formulated backward in time. Therefore, it should be suitable to change the temporal variable t in order to transform it into the initial-boundary value one. At first, setting $\hat{t}=T-t$ the time to maturity the option price and payoff functions transform into

$$u(S,M,\hat{t})=V(S,M,t), \qquad u_0(S,M)=\max(M-S,0) \quad (10)$$

and the pricing equation (5) leads to the forward parabolic partial differential equation (written in divergence form) for the unknown price function $u(S,M,\hat{t}):(0,\infty)^2\times(0,T)\to I\!R_0^+$ satisfying that

$$\frac{\partial u}{\partial\hat{t}}-\frac{\partial}{\partial S}\left(\frac{1}{2}\sigma^2S^2\frac{\partial u}{\partial S}\right)+\frac{\partial}{\partial S}\Big((\sigma^2-r)Su\Big)+(2r-\sigma^2)u=0, \quad S>0,\ M>0,\ 0\le\hat{t}<T. \quad (11)$$

Note that (11) is supplemented by jump conditions transformed in the similar manner as option value in (10).

Secondly, we restrict the problem (11) on a bounded (S,M)-domain and add suitable boundary conditions. For this purpose let S_{max} and M_{max} denote the maximal sufficient value of the underlying asset and maximal possible value of its maximum, respectively. Without loss of generality $M_{max}=S_{max}$, i.e., we consider the square domain $\Omega:=(0,S_{max})^2$ and the mixed boundary conditions are prescribed in the following way. Since the variable M is not present in the differential operator in (11), the convection does not propagate in the M-direction and thus no boundary condition has to be imposed on boundaries parallel to the S-axis. For the particular

situation $S = S_{max}$, we can argue that the value of the floating strike lookback put option should be insensitive to infinitesimal changes in M. On line $S = 0$ the put-call parity is simply used, i.e., the price of puts corresponds to the discounted value of the maximum M, because calls become worthless on this boundary. These boundary conditions are mathematically formulated as $\frac{\partial u}{\partial M}(S_{max}, M, \hat{t}) = 0$ and $u(0, M, \hat{t}) = Me^{-r\hat{t}}$.

Finally, note that the pricing equation (11) is closely related to the convection-diffusion equation, dominant component of which depends on a relationship between the Black-Scholes parameters σ and r. Since the analytical (semi) closed-form solution of (11) is not attainable with regard to jump conditions, one has to rely on the numerical schemes for solving of such equation taking into account its properties, see next section.

4 Numerical approach

We use a numerical approach based on the DG framework (see [10] for a complete overview), where the approximate solution is sought in the finite dimensional space S_h^p with dimension DOF and basis $\{v_{h,j}\}_{j=1}^{DOF}$ consisting from piecewise polynomial, generally discontinuous, functions of the p-th order defined on the domain Ω. Similarly as in [8], we introduce the semi-discrete solution $u_h = u_h(\hat{t})$ represented by the system of the ordinary differential equations supplemented by jump conditions at sampling dates $\{\hat{t}_i\}_{i=1}^{N}$ as

$$\frac{d}{d\hat{t}}\left(u_h, v_h\right) + \mathcal{A}_h(u_h, v_h) = l_h(v_h)(\hat{t}) \qquad \forall\, v_h \in S_h^p,\ a.a.\ \hat{t} \in (0,T), \tag{12}$$

$$\lim_{\varepsilon \to 0+} u_h(\hat{t} - \varepsilon) = \lim_{\varepsilon \to 0+} \mathcal{J}_{\mathcal{C}}\left(u_h(\hat{t} + \varepsilon)\right), \qquad \hat{t} \in \{\hat{t}_i\}_{i=1}^{N}, \tag{13}$$

where $u_h(0)$ is given by payoff u_0 from (10), $(\cdot,\cdot)$ denotes the inner product in $L^2(\Omega)$ and the bilinear form $\mathcal{A}_h(\cdot,\cdot)$ stands for the DG semi-discrete variant of diffusion, convection and reaction operator from (11) accompanied with penalties and stabilizations. Finally, the right-hand side form $l_h(\cdot)(\hat{t})$ contains terms arising from boundary conditions. The detailed description of the afore-mentioned forms can be found in [7].

Further, to obtain the fully discrete DG scheme, we discretize in temporal variable $\hat{t}$. We follow approach from [9], where we consider Crank-Nicolson scheme, which is practically unconditionally stable and gives the second order convergence in time. It is equivalently written as the weighted average of forward Euler and backward Euler methods. Let τ be the constant time step of partition $(0,T)$ capturing all sampling dates and $U_k = \{\beta_j^k\}_{j=1}^{DOF}$ be the vector of real coefficients in the linear combination of basis functions for the DG approximate solution of problem (12) at time $\hat{t}_k \in [0,T]$. Then at each time level we obtain the sparse matrix equation with discrete version of jump condition (13), i.e.,

$$\left(\mathbf{M} + \frac{\tau}{2}\mathbf{A}\right) U_{k+1} = \left(\mathbf{M} - \frac{\tau}{2}\mathbf{A}\right) U_k + \frac{\tau}{2}\left(F_{k+1} + F_k\right), \tag{14}$$

$$\sum_{j=1}^{DOF} \beta_j^k v_{h,j}(S,M) = \mathcal{J}_{\mathcal{C}}\left(\sum_{j=1}^{DOF} \beta_j^k v_{h,j}(S,M)\right), \qquad \hat{t}_k \in \{\hat{t}_i\}_{i=1}^{N},\ (S,M) \in \Omega, \tag{15}$$

where the vector U_k is related to the DG solution $u_h(\hat{t}_k)$, the matrix $\mathbf{M}$ to the mass matrix, the matrix $\mathbf{A}$ to the bilinear form $\mathcal{A}_h$ and the vector F_k to the boundary conditions, respectively. The resulting DG discrete solution arises from the unique solvability of the linear algebraic system (14).

For the sake of clarity, the whole numerical procedure for valuing of discretely sampled lookback options can be summarized into the following processes:

(I) Solve the Black-Scholes equation according to the scheme (14) for all time levels till the next sampling date $\hat{t}_i$ (or maturity T) using the value of the option at the previous sampling date as a new initial condition except the situation before the first sampling date when the original initial condition (10) is used.

(II) Apply the appropriate jump contiotion (15) across the current sampling date to set the value of the option immediately after the present sampling date.

(III) Repeat (I) and (II) as needed to reach the final time T.

5 Reference numerical experiments

In this section we give some numerical results for discretely sampled European lookback floating strike put options presented on reference benchmarks and compared to other commonly used numerical methods for different polynomial orders (linear, quadratic and cubic) of approximation as well as for different sets of sampling dates – cases (S1)–(S4) listed below. All computations are carried out with an algorithm implemented in the solver Freefem++, i.e., a mesh generation, the DG discretization, assembly of the corresponding linear algebraic problem, its solving and enforce the jumping conditions. The detailed description can be found in [6]. To specify the implementation

settings in each simulation the time step is set proportional to one business day as $\tau = 1/250$ and GMRES is used as a sparse solver for (14).

As the numerical experiment we consider the reference model problem from [11] with the following parameters: $T = 1.0$, $r = 0.1$, $\sigma = 0.2$, $S_{max} = M_{max} = 400$, $M_{ref} = 200$, where M_{ref} determines the current maximum. For the purpose of a broader comparison the several sampling cases are investigated, namely

Case (S1): sampling at times $0.5, 1.5, 2.5, \ldots, 10.5, 11.5$ months;
Case (S2): sampling at times $1.5, 3.5, 5.5, 7.5, 9.5, 11.5$ months;
Case (S3): sampling at times $3.5, 7.5, 11.5$ months;
Case (S4): no sampling.

Note that the case (S4) is exactly equivalent to a simple vanilla put option with a strike price equal to M_{ref}.

First, we compare the DG method used in this work with the finite difference ones from [11], [12] and the finite element one presented in [5], for one particular case of sampling (S1). The comparative results evaluated at three reference asset values $S_{ref} \in \{180, 200, 220\}$ at one year maturity are presented in Table 2. Actually, the results given in this table are not option prices but their scaled values by the factor M_{ref}. One can easily observed that DG approach for piecewise linear, quadratic as well as cubic approximations gives fairly the same results as the other methods. This fact is also illustrated in Figure 1 (left).

S_{ref}	$\frac{S_{ref}}{M_{ref}}$	ref. value [11]	ref. value [12]	ref. value [5]	DG(P_1)	DG(P_2)	DG(P_3)
			$n_x = 389$	$n_x = 400$	$n_x = 400$	$n_x = 200$	$n_x = 140$
180	0.9	0.101	0.10025	0.10035	0.100125	0.100148	0.100157
200	1.0	0.089	0.08885	0.08887	0.088608	0.088637	0.088635
220	1.1	0.095	0.09546	0.09502	0.095237	0.095262	0.095247

Table 2 Comparison of scaled option values evaluated at reference node $[S_{ref}, M_{ref}]$ for sampling strategy (S1), where n_x denotes the number of equidistant space points in the S-direction

Further, we investigate the behaviour of lookback option values with respect to different lengths of sampling interval, see Table 3. The computed piecewise linear option values for all sampling strategies are in quite good agreement with empirical findings as well as with results from [11] obtained by the finite difference scheme. It is apparent that option prices decrease as the length of the sampling interval prolongs, i.e., as the number of sampling dates decreases from (S1) to (S4). In other words, increasing the frequency of measurement of the maximum increases the option value. This typical behaviour of lookback options is also depicted in Figure 1 (right).

S_{ref}	$\frac{S_{ref}}{M_{ref}}$	Case (S1)		Case (S2)		Case (S3)		Case (S4)	
		FD [11]	DG(P_1)	FD [11]	DG(P_1)	FD [11]	DG(P_1)	FD [11]	DG(P_1)
180	0.9	0.101	0.100125	0.094	0.093586	0.087	0.086074	0.074	0.074331
200	1.0	0.089	0.088608	0.079	0.078130	0.067	0.065333	0.038	0.037517
220	1.1	0.095	0.095237	0.083	0.083136	0.068	0.068223	0.017	0.017308

Table 3 Comparison of scaled option values evaluated at reference node $[S_{ref}, M_{ref}]$ in dependence on sampling strategies considered

Finally, let us say that values of floating strike lookback put options reach their minimum in the neighbourhood of the point where the asset price is equal to the current maximum (i.e., $S_{ref} = M_{ref}$). This is not true for sampling strategy (S4), which corresponds to the simple vanilla put option having its minimum asymptotically as asset price tends to infinity, see Figure 1.

6 Conclusion

Pricing of options is very challenging and no less important part of financial engineering. In this paper we have presented numerical scheme based on DG approach for pricing of special types of path dependent options, namely discrete lookback put with floating strike. Comparison of the results with reference values presented in [11], [12] and [5] does not show any significant differences. However, the advantages of DG approach for lookback option pricing can be seen in that using the DG technique also for the temporal coordinates enables us to describe discrete sampling of the maximum or minimum of the underlying asset more accurately. Another benefit of DG approximations lies in its easier extension to other classes of path-dependent options with more complex payoff functions.

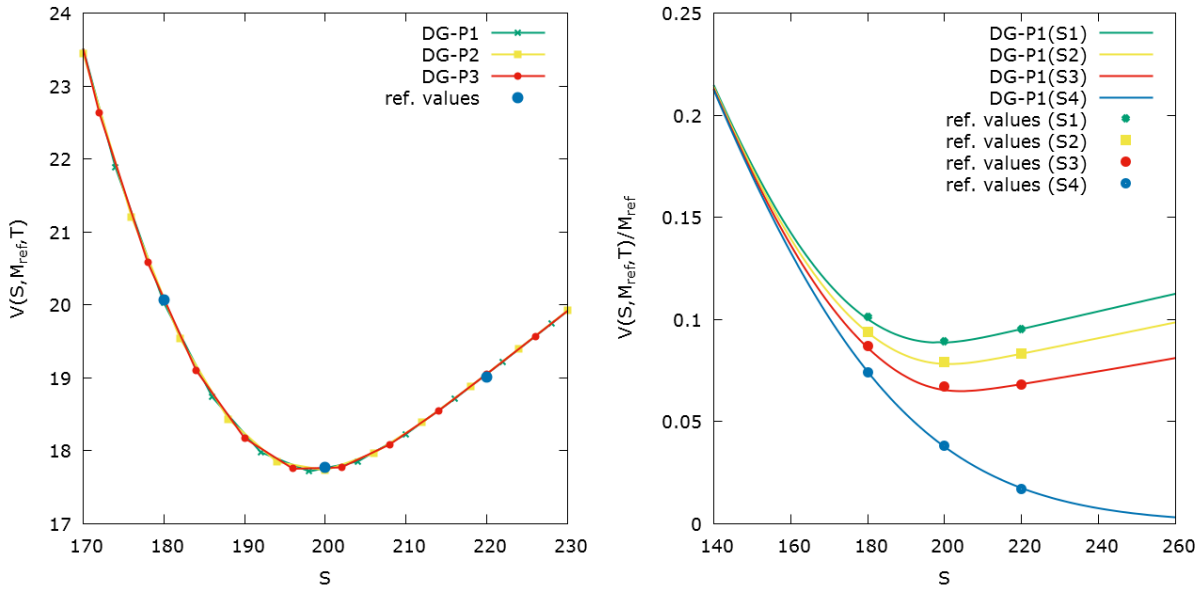


Figure 1 Discrete non-scaled option values for piecewise linear, quadratic a cubic approximations and (S1) sampling strategy (left), discrete piecewise linear scaled option values for (S1)-(S4) sampling strategies (right)

Acknowledgements

Both authors were supported through the Czech Science Foundation (GAČR) under project 16-09541S. The support is greatly acknowledged. Furthermore, the second author also acknowledges the support provided within SP2017/32, an SGS research project of VŠB-TU Ostrava.

References

[1] Achdou, Y., and Pironneau, O.: *Computational Methods for Option Pricing*. Society for Industrial and Applied Mathematics, Philadelphia, 2005.

[2] Black, F., and Scholes, M.: The pricing of options and corporate liabilities, *Journal of Political Economy* **81** (1973), 637–659.

[3] Černá, D., and Finěk, V.: On a sparse representation of a n-dimensional Laplacian in wavelet coordinates, *Results in Mathematics* **69** (2016), 225–243.

[4] Černá, D., and Finěk, V.: Sparse wavelet representation of differential operators with piecewise polynomial coefficients, *Axioms* **6** (2017), no. 4.

[5] Foufas, G., and Larson, M.G.: *Valuing European, barrier, and lookback options using the finite element method and duality techniques*. Chalmers University of Technology, Sweden, 2004.

[6] Hecht, F.: New development in FreeFem++, *Journal of Numerical Mathematics* **20** (2012), no. 3-4, 251–265.

[7] Hozman, J.: Discontinous Galerkin method for the numerical solution of option pricing, *Aplimat – Journal of Applied Mathematics* **5** (2012), no. 2, 271-280.

[8] Hozman, J., and Tichý, T.: DG method for numerical pricing of multi-asset Asian options—The case of options with floating strike, *Applications of Mathematics* **62** (2017), no. 2, 171-195.

[9] Hozman, J., and Tichý, T.: DG solver for one-factor and two-factor Black-Scholes models. In: *Proceedings of the 8th International Scientific Conference Managing and Modelling of Financial Risks* (M. Čulík, eds.), VŠB-TU Ostrava, Ostrava, 2015, 323-332.

[10] Riviére, B.: *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*. Society for Industrial and Applied Mathematics, Philadelphia, 2008.

[11] Willmott, P., and Dewynne, J., and Howison, J.: *Option pricing: mathematical models and computations*. Oxford financial press, 1993.

[12] Zvan, R., and Vetzal, K.R., and Forsyth, P.A.: PDE methods for pricing barrier options, *Journal of Economic Dynamics and Control* **24** (2000), no. 11-12, 1563–1590.

Möbius.
Environment for learning mathematical modelling in economy

Jiří Hřebíček [1]

Abstract. Möbius™ is a comprehensive online courseware environment that focuses on science, technology, engineering, and mathematics (STEM). Möbius is built on the notion that people learn by doing. Students can explore important concepts using engaging, interactive applications, visualize problems and solutions, and test their understanding by answering questions that are graded instantly. Throughout the entire lesson, students remain actively engaged with the material and receive constant feedback that solidifies their understanding.
Möbius is designed to help academic institutions move their STEM courses online, providing a rich development and deployment environment that meets the particular challenges of mathematically based courses in economy and mathematical modelling.

Keywords: STEM, Möbius system, Maple, learning, mathematical modelling.

JEL Classification: A20
AMS Classification: 97Bxx, 97Dxx

1 Introduction

Online courses of natural, social and technical sciences have become a key part of the instructional offerings of most learning institutions, from prestigious universities to local high schools. They allow students greater scheduling flexibility, reach a broader geographic pool, and reduce costs. Universities today have to offer online courses and supplementary resources to stay competitive. Online courses frequently offer a mix of text and videos, chat rooms, and forum discussions with mandatory participation. The learning process, in *Science, Technology, Engineering*, and *Mathematics* (STEM) subjects [1], means planning, managing, and providing scientific research, professional and technical services (e.g., physical science, social science, engineering) including laboratory and testing services, and research and development services. It includes occupations such as engineer (all types), technician (all types), drafter, chemist, mathematician, archaeologist, biologist metallurgist, math or science teacher, nutritionist, oceanographer, materials analyst, computer programmer, technical sales manager, CAD operator, economist, and technical writer. STEM education is an interdisciplinary approach to learning where rigorous academic concepts are coupled with real-world lessons as students apply science, technology, engineering, and mathematics in contexts that make connections between school, community, work, and the global enterprise enabling the development of STEM literacy and with it the ability to compete in the new economy [2].

European Schoolnet (http://www.eun.org/focus-areas/stem) is at the forefront of the debate on how to attract more people to science and technology to address the future skills gap that Europe is facing. STEM is one of European Schoolnet's major thematic domains. It has been involved in more than thirty STEM education initiatives, financed through European Schoolnet members, industry partners, or by European Union's funding programmes [3].

The understanding and scope of STEM skills varies widely from country to country [3]. Supply is relatively clearly identified in terms of qualifications achieved in STEM subjects, although definitions of STEM subjects can vary. For example, medicine, structural engineering and sports science are not included in some definitions. Basic STEM subjects typically include: *Mathematics; Chemistry; Computer Science; Biology; Physics; Architecture; Econometrics.* Further *General, Civil, Electrical, Electronics, Communications, Mechanical and Chemical Engineering*.

While economics is a social science because it studies social behaviours, its heavy utilization of mathematics, quantitative reasoning, and technology suggests that economics fits within the STEM classification. A typical bachelor study in economics will require taking *calculus I, II, III, linear algebra, differential equations, probability theory, statistics, numerical analysis, and modelling*. Ambitious and talentwise students often take additional mathematics and computer science courses, enabling them to act in roles like data scientists.

Online course environments are only suitable for STEM course if they can handle mathematics-based content. Here are just a few of the challenges that must be addressed:

[1] Masaryk University, Institute of Biostatiscs and Analyses, Kamenice 126/3, 625 00 Brno, hrebicek@iba.muni.cz..

- The *mathematics for economics* needs to "*look right*", whether it is displayed in the text or the student enters it as the response to a question.
- The STEM system must can display a *wide variety of 2-D and 3-D plots*, preferably that change dynamically, preferably without requiring intensive work on the part of the authors.
- *Assessment questions* in STEM system must be free-response. Students must be able to show they have worked out the correct solution, not simply picked the correct answer out of a list.
- *Automatic grading* in STEM system must deal with mathematical equivalency.
- The assessment engine in STEM system should be able to handle a wide variety of mathematics, so it can handle a wide variety of STEM courses.

On 19 April 2017, the Canadian company Maplesoft™ (www.maplesoft.com) announced the public release of Möbius system [5], the online courseware environment that focuses on STEM education. After months of extensive pilot testing at select leading academic institutions around the world, Möbius is now available to all institutions for their online learning needs. Customers will also soon have the option of adopting Möbius-based course materials developed by the University of Waterloo (http://courseware.cemc.uwaterloo.ca/), a leading institution in STEM education.

Möbius is built on the notion that people learn by doing. With Möbius, students can *explore important concepts* using *engaging, interactive applications, visualize problems and solutions, and test their understanding by answering questions that are graded instantly*. Throughout the entire lesson, students remain actively engaged with the material and receive constant feedback that solidifies their understanding. In other words, they have to pay attention.

The Möbius system consists of the following basic components [5]: *Question Repository* - collection of questions. *Assignments* - tests containing questions from the question repository with policies that control their form and availability. *Gradebook* - results of student assignments. *Slideshows* - create engaging visual presentations of courses (lessons) with embedded video and interactive content. *Lessons* - organize visual presentations, text, Mathematical Applications, questions, and assignments. *Calendar* - create and organize scheduled events (examination, tests etc.)

There are four roles of users in Möbius [5]: *System Administrator, Instructor, Proctor,* and *Student*. Generally, System Administrator (SA) creates user accounts. Teachers can have role of *Instructor* or *Proctor*. The SA can give instructors the ability to create user accounts (e.g. proctors and students).

The paper introduces roles of users of Möbius components in the second chapter and Möbius components in the third chapter. Möbius course for mathematical modelling in economy will be demonstrated in the presentation of the paper at conference.

2 Möbius user roles

Each user must have an account in the system Möbius. All users log into Möbius through the same login screen using their user name and password. A default role of user is assigned by the *System administrator*. Users can have different roles in different classes. After users are logged in, they have access to different *Class Homepages* or menu items (*Class User Manager, Proctor Tools, Content Repository, Gradebook* and *External*) [5] depending on their role in that class. *Class User Manager* enables [5]: *Search for Users; Show Active Users; Import Users From Roster; Remove Users Using Roster; Register Users*; and *Remove Users*. *Proctor Tools* enable manage proctors and authorization of assignments. For example, an instructor can be a student in one class, with student privileges in that class, and he or she will have instructor privileges in the class he or she is instructing.

2.1 System Administrator

Generally, the *System Administrator* (SA) creates all user accounts. The SA can give *Instructors* the ability to create *Proctor* and *Student* accounts. Users can have different roles in different classes. The SA or instructor provides students with a username and password to log in.

2.2 Instructor

The first step of instructor is to log into the Möbius system using his username and password provided by the SA or *course coordinator*. Logging into the Möbius system brings *Instructor* to the *System Homepage* (Figure 1).

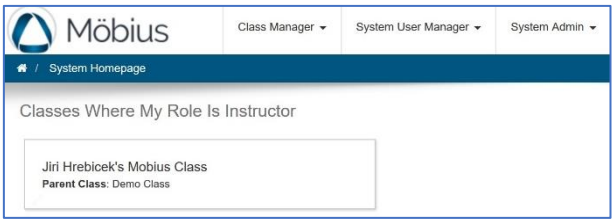


Figure 1 System Homepage

After instructor is logged into the system, he can create a *new class*. He can upload a student roster or set open registration so that users can enrol themselves in his class. Any created classes are listed next on the *System Homepage*. Additionally, any classes in which users are a student or proctor are also listed, under their own sections. From the *System Homepage*, user can click in menu on a class name link to open the *Class Homepage*.

2.3 Proctors

There are two types of proctors: *Global* and *Local: Global proctors* are defined by the SA and can give authorization for any class; *Local proctors* are defined by the *Instructor* and can only give authorization for a particular class. Proctors can give authorization directly at the student's computer, or remotely through item *Proctor Tools* in menu. Typically, a proctor is physically present in the same computer laboratory during testing sessions. When notified by the student, proctors can also conveniently authorize a student examination session from their own workstation using Proctor Tools. Proctors can then notify the student(s) they are giving permission to.

2.4 Students

Generally, the SA creates *Student* accounts in system. The SA can give instructors the ability to create Student accounts. Student ID must be at least one alphanumeric character and must be unique. To register students in instructor course, they must have an account in the system.

3 Möbius system components

All users log into Möbius through the same login screen using their user name and password. After users are logged in, they have access to *Class Homepages* connected with course or menu items depending on their role (instructor, proctor, student) in that class (Figure 2).

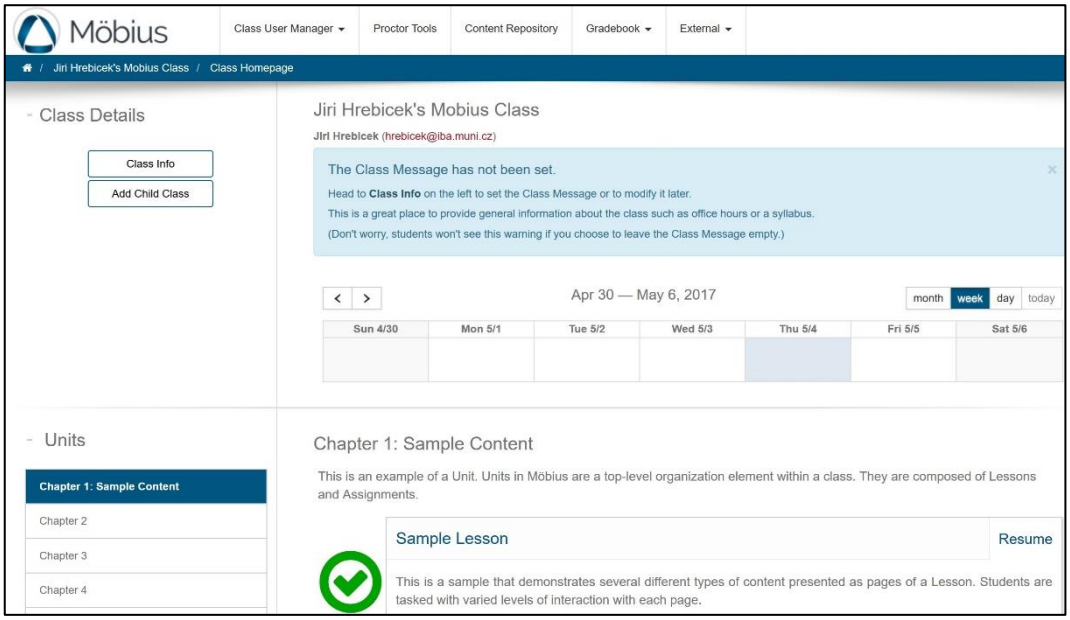


Figure 2 Class Homepage

3.1 Creating and Managing Classes

Teachers (instructors) begin with creating a new class. Within their classes, they have access to assignments, resource files, the *Content Repository*, and the *Gradebook* for the class. The SA may have already created a class

for teacher (instructor). After instructor logins, he can access his class by clicking the link under the *Classes* in menu where his role is *Instructor*. Instructor can use one of the following options to create or share his class:

- *Create a New, Parent Class*. He can create a new class from the *System Homepage*, select the appropriate *Class Manager* menu.
- *Create a Child Class*. A child class inherits content from a *parent class*, such as questions and pre-populated assignments. To create a new class that inherits content from a featured class that instructor uses System Homepage to select it form the Class Manager menu (see Figure 2).
- *Share class*. When instructor wants to share a class, he must create a *single parent class* whose content and assignments are distributed to child classes. The content and assignments are inherited by each child class when a new class is created and the Inherit content from option is selected in the class registration form.

After a class is set up, users can access assignments, quizzes, proctored tests, and results through the *Class Homepage*. Students can register in instructor's class if he allows for open enrolment; otherwise he can add students by uploading a student roster or through the Class User Manager.

3.2 Proctor Tools

Instructor can add a proctor to his class to ensure security in exam settings. Proctors can log in to their own computer and perform proctoring duties using the *Proctor Tools* utility from the *Class Homepage* (Figure 2). Proctors can authorize students to start a test as well as authorize grading of a test.

3.3 Content Repository

Instructors creates *self-scoring assignments* with *questions* drawn from the *Content Repository* (CR), see Figure 3. The system's assignments can help automate common learning activities that involve questions and problems in classes such as practice and proctored examinations, tutorial sessions, homework assignments, and quizzes. The system can report on individual student performance, performance across a class section or across multiple sections of a class, and a statistical item analysis of questions. The system can be configured to enable easy sharing of class content across sections of large classes.

The CR is the central location for managing *questions*. Teachers can use item *Question/Text* in menu *Create New* in the CR (Figure 3) to: *Create questions; Preview a question; View question details, including author and revisions; Search for questions; Organize questions; Share questions; Import questions; Export questions*.

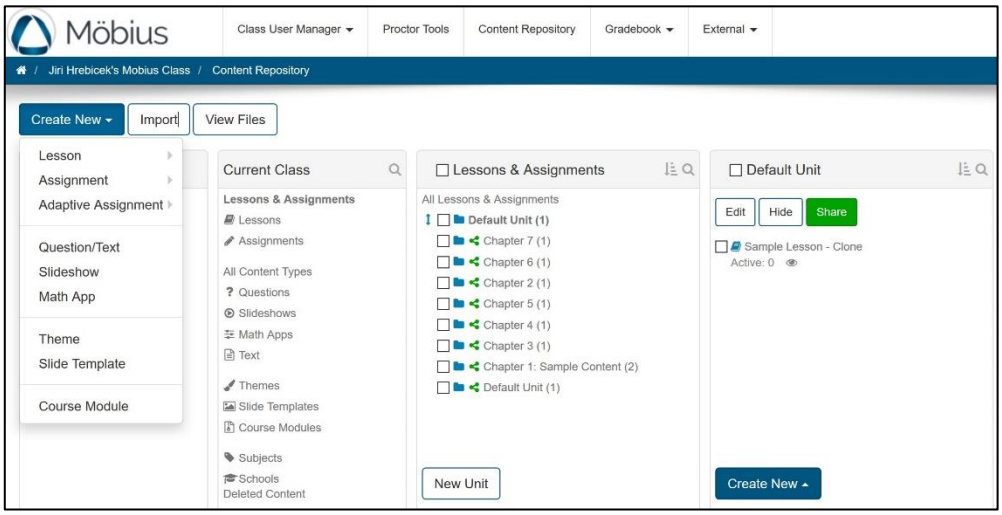


Figure 3 Content Repository

Teachers can access their personal Maple library archives (Maple Repository) [6] from within their Maple questions. To do this, they must first upload their archive files into the class file space on the server. When they create a Maple-graded questions, they then specify the location of the archive as part of the question definition.

Lessons

Lesson item in menu *Create New* (Figure 3) is a way to offer activity-based learning. Teachers (instructors or proctors) can present content, ask leading questions, and offer interactive applications for exploratory learning. The *Student* learns by engaging.

A lesson consists of a collection of pages. Each page consists of blocks. Blocks can be questions, text, Mathematical Applications (Math Apps), slideshows, or question groups. Pages provide an easy way to organize lesson content. For example, teacher may want to create separate blocks to split up lecture presentations by subtopic. He may also decide to create separate blocks for presentations, quiz questions, and solutions to quiz questions. The group of blocks can be encapsulated into course modules. These allow teachers (instructors or proctors) to share their class content with other teachers, including related questions, assignments, text and themes from the CR. After teachers create and export a course module, they can send it to other instructors who can upload it into their class and use it in its current state, or alter it to suit their needs. Course modules encapsulate from the source class and allow teachers to reinstall questions and assignments in the destination class [5]. After designing the lesson, they set the properties for the lesson. Creating lessons is done in the *Lesson Editor* that enables [5]: *Create new lessons; Edit existing lessons; Import questions, text, Math Apps, slideshows, and question groups into a lesson; Schedule availability and prerequisites of a lesson*.

Assignments

From the CR, teachers can also create *Assignments* for a range of purposes, including [5]: *Self-study practice; Mastery; Homework; Quizzes; Proctored exams requiring proctor validation*.

The Möbius obtains the *Assignment Editor* that allows instructor to create new assignments, edit the content, properties, or appearance of existing assignments [5]. Each student viewing instructor's assignment can potentially complete a unique set of questions. When an assignment is created, teachers specify the rules and policies for assignment access and due dates, grading parameters, and feedback options. After determining the content, rules, and policies, you publish an assignment to the class web site for students to use. Students can review the results of their previously completed assignments in the *Gradebook*.

Slideshow

Möbius slideshows provide the capability to embed images, text, math and interactive content, such as Mathematical Applications of Maple of MapleCloud and questions that require student response into your presentation. Slideshows can be set to run automatically, or to progress through each slide manually. All work with slideshows is accomplished through the *Slideshow Designer* menu. Teachers can use the Slideshow Designer to [5]: Create slideshows consisting of text, math, images, questions and Mathematical Applications of Maple; Record audio or upload an existing audio file for individual slides; Download and save recorded audio to your local machine; Add transitions and pauses corresponding to different elements on slide before, during, and after recording audio; Add algorithms and slideshow-level custom CSS (themes).

A key method of presenting course content in Möbius is the *Active Slideshow*. The Active Slideshow is like a great classroom lecture: a methodical presentation of information in an order that makes sense, combines a variety of learning elements with frequent "*checking in*" with the students to ensure understanding, and provides additional help when it is required. And unlike a traditional classroom, the experience is tailored to each student.

3.4 Gradebook

Gradebook menu consists of [5]: *Class Grades, Item Statistics, Grade Reports, Rubric Tables items*. After students complete an assignment, teacher can view the results in Gradebook or create downloadable performance reports that can be viewed in third-party gradebook programs or Microsoft® Excel. With Möbius, teachers (instructors) can generate comprehensive performance reports for individual students, individual class sections, or multiple class sections. They can also perform a statistical item analysis of questions.

The Gradebook has primary functions [5]: Viewing and analysing scores and statistics for students, question items assignments and lessons; Reviewing student results; Creating customized grade reports that reflect your grading policy. These grade reports can be saved or published for student viewing; Exporting grades to a file; Using customized rubrics to assign grades to students.

Teachers can use the Gradebook to view grade data organized by student, by question, by assignments or lessons. Grade reports allow teachers to [5]: Group assignments; Set criteria within an assignment group, such as "*drop the lowest grade*" and whether to ignore non-attempts or mark as zero; Designate weighting for each assignment in a group; Designate weighting for each assignment group in the overall grade; Designate an assignment group as extra credit; Save the grade report.

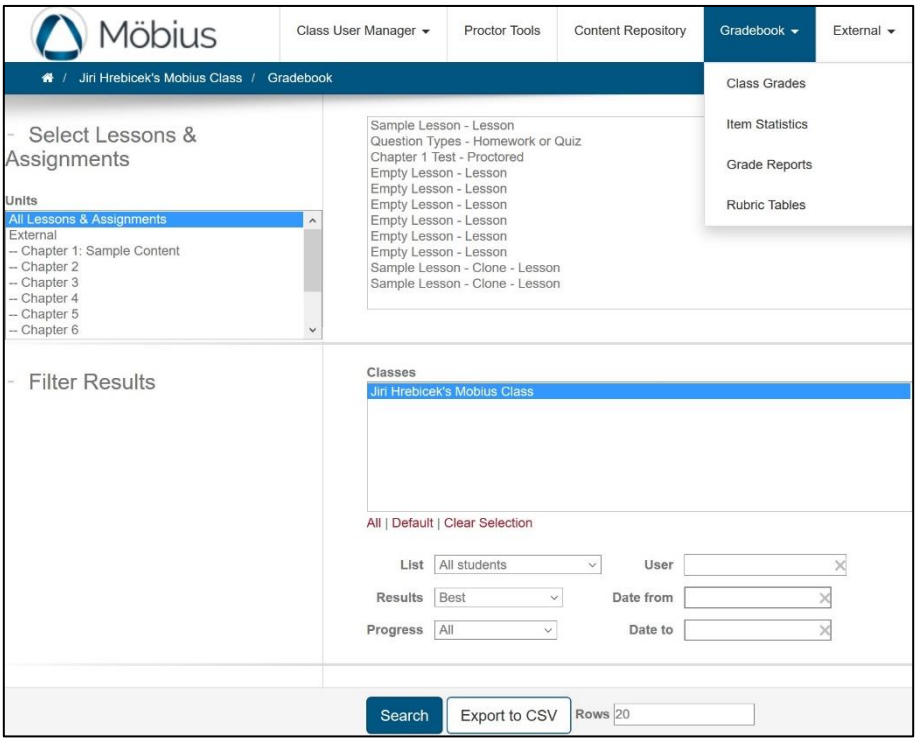


Figure 4 Gradebook

4 Conclusion

Möbius course content can include text, graphs, animations, videos, interactive applications, and formative and summative assessments. These elements can also be combined in a slideshow, so they only appear when the teacher (instructor) wants them too. In this way, the instructor can control the flow of information to their class, preventing students from jumping ahead before they are ready. In addition, the system Möbius:

- Reliably monitors student understanding throughout the entire lesson, so students are confident when they should be and find out immediately when they've gone off-track.
- Offers a variety of experiences in a single lesson, all of which are directly relevant to the materials being taught.
- Asks each student questions, and provides immediate feedback and even the opportunity to try additional questions if the student would benefit from the extra practice.

Möbius provides the tools which university teachers need to author and deliver rich, engaging online offerings for STEM education of mathematical modelling in economy. Möbius is designed to help academic institutions move their STEM courses online, providing a rich development and deployment environment that meets the challenges of mathematically based courses in economy and mathematical modelling.

Acknowledgements

This paper was supported by the Czech Science Foundation. Name of the Project:" Development of new methods of solving dynamic models of corporate processes management" GA16-03796S.

References

[1] Gonzalez, H.B., and Kuenzi, J.F.: Science, Technology, Engineering, and Mathematics (STEM) Education: A Primer. 2012. https://fas.org/sgp/crs/misc/R42642.pdf (last accessed 05.05.17.).
[2] Kelley, T.R., and Knowles, G.F.: A conceptual framework for integrated STEM education. *International Journal of STEM Education* (2016), DOI: 10.1186/s40594-016-0046-z.
[3] EU skills panorama 2014. April 2015. Analytical highlight focus on science, technology, engineering and mathematics (STEM) skills. http://skillspanorama.cedefop.europa.eu/sites/default/files/EUSP_AH_STEM_0.pdf (last accessed 05.05.17.).
[4] Möbius Online Help. http://www.maplesoft.com/support/help/Mobius/ (last accessed 05.05.17.).
[5] Mobius Instructor Help. Maplesoft, Waterloo, 2017.
[6] Maple Help System. http://www.maplesoft.com/support/help/Maple/view.aspx?path=worksheet/help/helpnavigator (last accessed 05.05.17.).

Application of Kohonen SOM Learning in Crisis Prediction

Radek Hrebik[1], Jaromir Kukal[2]

Abstract. The Self-Organized Mapping (SOM) is a traditional tool for multidimensional data analysis which overperforms analytical power of cluster analysis. But there are possible difficulties when the SOM is applied to data patterns of large size. Our approach macro-economical data analysis is based on logarithmic differences, pattern dimensionality reduction and finalization of data analysis using Kohonen SOM learning. This general methodology was applied to the statistic data describing the economic situation of thirty five countries during more than twenty years. The regularly published data come from statistics of European Commission. The aim is to identify similar groups of countries and characterized the similarity. The role of SOM topology, learning strategy and reduced pattern size can be also used to crisis prediction based on similarities with countries already suffering with crisis.

Keywords: SOM, Kohonen learning, artificial neural network, macroeconomic indicators, crisis prediction.

JEL classification: C44
AMS classification: 90C15

1 Introduction

In our research we deal with basic economical indicators which are published on regular basis. The Self-Organized Mapping (SOM) represents a traditional tool for multidimensional data analysis which overperforms analytical power of cluster analysis. We face possible difficulties applying the SOM to data patterns of large size. So we have to make data preprocessing. Our approach of macroeconomic data analysis is based on logarithmic differences, pattern dimensionality reduction and finalization of data analysis using Kohonen SOM learning.

This general methodology was applied to the statistic data describing the economic situation of more than thirty countries during more than twenty years. The regularly published data come from statistics of European Commission. The aim is to identify similar groups of countries and characterize the similarity. The role of SOM topology, learning strategy and reduced pattern size can be also used to crisis prediction based on similarities with countries already suffering with crisis.

2 Kohonen Learning

Kohonen Self Organized Map (SOM) is organized as follows. Let $m, n, H \in \mathbb{N}$ be number of patterns, pattern dimensionality and number of SOM neurons [2]. The individual patterns are $\mathbf{x}_j \in \mathbb{R}^n$ where $j = 1, ..., m$ and form the pattern set $\mathcal{S} = \{\mathbf{x}_1, ..., \mathbf{x}_m\}$. The topology of SOM [6] is described by undirected graph $\mathcal{G}$ of H vertices which are connected with unit length edges. The SOM topology matrix $\mathcal{G} \in \{0,1\}^{H\times H}$ generates mutual vertex distances $\Delta_{i,j}$ for $1 \le i, j \le H$. The result of SOM learning is the system of weights [8] $\mathbf{w}_i \in \mathbb{R}^n$ where , $i = 1, ..., H$. We begins with random weights setting $\mathbf{w}_i(0)$. The weights evolve during learning process and their values are denoted as $\mathbf{w}_i(q)$ where $q \in \mathbb{N}_0$.

Kohenen learning rules [5] are very simple. The weight of i-th neuron is changed in q-th step by rule

$$\mathbf{w}_i(q) = \mathbf{w}_i(q-1) + \alpha(q) \cdot c_{i,q} \cdot (\mathbf{x}_q - \mathbf{w}_i(q-1)) \tag{1}$$

for $i = 1, ..., H$, $\mathbf{x}_q \sim \mathrm{U}(\mathcal{S})$ is uniformly selected pattern from $\mathcal{S}$, $c_{i,q}$ is space factor and $\alpha(q) > 0$ is ageing function which is supposed to be non-increasing. The winner is also selected according to Kohonen rule [5] as

$$\varphi_q \in \underset{k=1,...,H}{\arg\min} \|\mathbf{x}_q - \mathbf{w}_k\|_2. \tag{2}$$

[1]FNSPE, CTU in Prague, Trojanova 13, Prague 2, Czech Republic, Radek.Hrebik@seznam.cz
[2]FNSPE, CTU in Prague, Trojanova 13, Prague 2, Czech Republic, Jaromir.Kukal@fjfi.cvut.cz

We recommend generate the initial weights from the multi-varietal Gaussian distribution as

$$\mathbf{w}_i(0) \sim \mathrm{N}(\mathrm{E}\mathbf{X}, \mathrm{var}\mathbf{X}/100) \tag{3}$$

for $i = 1,...,H$. The space factor $c_{i,q}$ is calculated using mutual vertex distances as follows. Using learning radius $R_q > 0$ and index of winner vertex ϕ_q, we directly evaluate

$$c_{i,q} = \exp\left(-\frac{\Delta^2_{i,\phi_q}}{2R_q^2}\right)$$

according to Gaussian decay. The final learning strategy consists of $E \in \mathbb{N}$ learning epoch which we characterized by triplets (α_k, R_k, N_k) for $k = 1,...,E$. Here, α_k is ageing factor, R_k is learning radius, and N_k is number of learning steps in k-th epoch.

3 Quality Measures

The basic way of quality measurement design is based on measuring distances. The Euclidean distance of points $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$ is denoted $\mathrm{d}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$.

Using the pattern $\mathbf{x}_j$ we can investigate the distances to weights $\mathbf{w}_k$ and define winner as

$$\mathrm{win}(j) \in \underset{k=1,...,H}{\arg\min}\, \mathrm{d}(\mathbf{x}_j - \mathbf{w}_k) \tag{4}$$

but the function win(j) is of stochastic nature due to possible distance equities. In some cases we found the winner but one i. e. the second winner which is defined as

$$\mathrm{win2}(j) \in \underset{k\in\mathcal{M}_j}{\arg\min}\, \mathrm{d}(\mathbf{x}_j - \mathbf{w}_k) \tag{5}$$

where $\mathcal{M}_j = \{1,...,H\} \setminus \{\mathrm{win}(j)\}$.

Using distances and winners we can design traditional measures of various nature.

3.1 Distance penalization

The Quantization Error (QE) is traditionally related to all forms of vector quantization and clustering algorithms [7]. Using linear penalisation we directly penalise the distances between patterns and corresponding winner weights as

$$QE_1 = \sum_{j=1}^{m} \mathrm{d}(\mathbf{x}_j, \mathbf{w}_{\mathrm{win}(j)}). \tag{6}$$

The quadratic penalisation

$$QE_2 = \sum_{j=1}^{m} \mathrm{d}^2(\mathbf{x}_j, \mathbf{w}_{\mathrm{win}(j)}) \tag{7}$$

is also frequently used but has higher sensitivity to outliers.

3.2 Topographic error

General topographic rule is: if two objects are close in reality they must be closed also in the map. Using this principle the Topographic error (TE) [3] is defined as

$$TE = 1 - \frac{1}{m}\sum_{j=1}^{m} g_{\mathrm{win}(j),\mathrm{win2}(j)} \tag{8}$$

where $\mathbb{G} \in \{0,1\}^{H\times H}$ is SOM topology matrix with $g_{u,v} = \mathrm{I}(\|\mathbf{p}_v - \mathbf{p}_v\|_2 \leq 1)$. The main advantage of TE is in its robustness to outliers. Therefore we use this criterion as main quality measure in this study.

3.3 Correlation based measures

The correlations between mutual distances of patterns and mutual distances of winner weights can be directly used as quality measures.

Let i, j be pattern indexes. The mutual pattern distances can be defined as $d_{i,j} = \mathrm{d}(\mathbf{x}_i, \mathbf{x}_j)$. The mutual distances of corresponding weights are $\delta_{i,j} = \mathrm{d}(\mathbf{w}_{win(i)}, \mathbf{w}_{win(j)})$.

Finally, we obtain $m(m-1)/2$ pairs of corresponding distances and directly calculate Pearson correlation coefficient r, Spearmann ρ or Kendall τ coefficient as quality measure.

3.4 Time Complexity of Measures

The evaluations of QE_1, QE_2 and TE are very fast with time complexity $\mathrm{O}(mnH)$. The evaluation of correlation measures is more complex. The Pearson r has time complexity $\mathrm{O}(mnH + m^2)$ due to simple statistics over $m(m-1)/2$ distance pairs. The Spearmann ρ is complicated with pair sorting and its time complexity is $\mathrm{O}(mnH + m^2 \log(m))$. The Kendall τ is not recommended for large pattern sets due to time complexity $\mathrm{O}(mnH + m^4)$.

4 Case Study: Economical Indicators

As input data we used the main economic indicators. Data has been selected from Statistical Annex of European Economy presented by European Commission in autumn 2016 [1]. As analysis input serve the thirty five countries from the whole world, majority are the European countries. The indicators are observed in years 1993 to 2016. Selected indicators are the total population, unemployment rate, gross domestic product at current market prices, private final consumption expenditure at current prices, gross fixed capital formation at current prices, domestic demand including stocks, exports of goods and services, imports of goods and services and gross national saving. Nine indicators are monitored in total. The main aim of our research is based on data for each country.

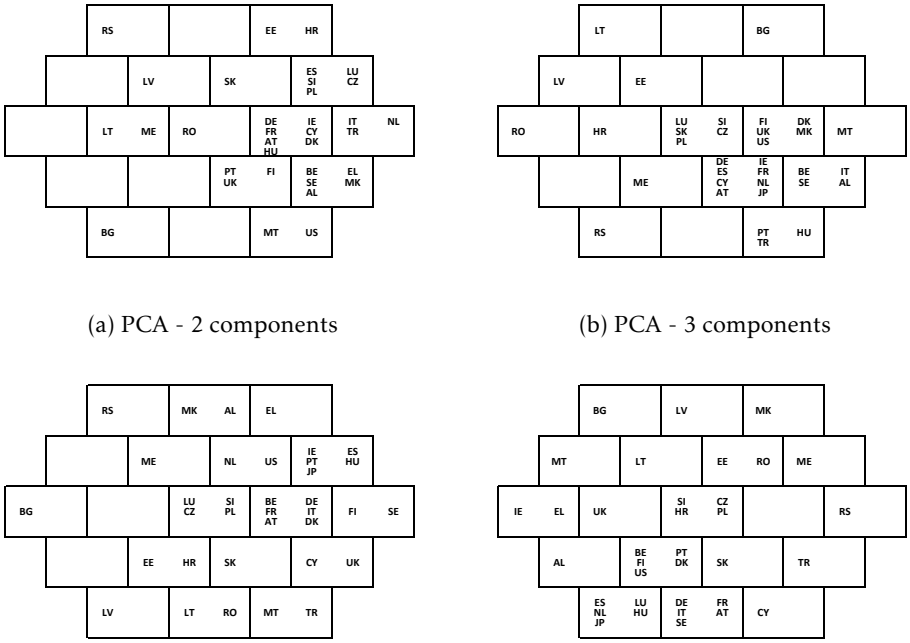
As the dimensionality of input data is quite high, represented by main nine indicators in each year, we use principal component analysis for data dimension reduction. We prefer the standardize variant of PCA which divides the components into square roots of adequate eigenvalues. This approach is frequently called data whitening. The main advantage of the standardization is in identity covariance matrix which generates the components in unified form. We studied data whitening for $D = 2, 3, 4, 5$. Then we applied Kohonen SOM with hexagonal topology with node number $H = 7, 19$. The SOM learning with Gaussian decay was driven by two strategies. For $H = 7$ we used only $E = 2$ with $\alpha = (0.1, 0.05)$, $R = (2, 1)$, $N_k = 1000$. The larger SOM with $H = 19$ was learned for $E = 9$ with $\alpha = (0.1, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02, 0.01)$, $R = (5, 3, 3, 1.5, 1, 0.7, 0.5, 0.3, 0.2)$ and $N_k = 1000$. Our aim was to obtain the SOM with zero topographical error (TE) and minimum possible quadratic penalisation (QE_1). The results of QE_1 are captured in table 1.

Table 1: Optimal QE_1 measures

D	SOM_7	SOM_{19}
2	0.002	0.001
3	0.003	0.002
4	0.010	0.007
5	0.020	0.010

5 Results

In all cases we obtained zero values of TE which means that learning was executed well. It is evident from table 1 that SOM_{19} generates results with lower value of QE_1 which is rising with growing dimension. The distribution of countries is captured in figure 1. We see the PCA with 2 components as the best solution and resulting SOM. The different groups of countries were identified. They tell us about the similarities of the concrete countries. The main thing what we can see is the position of Germany, which is usually in the same group as France. In the case of Czech Republic its position depends on number of components but we are in the same group with Poland and Slovenia in all cases. In all cases



(a) PCA - 2 components

(b) PCA - 3 components

(c) PCA - 4 components

(d) PCA - 5 components

Figure 1: Results for $H = 19$ and different number of components

there are relative compact group of traditional countries which slightly differs each other. The positions of countries with extreme macro-economical behaviours differ with whitening dimensionality. The results are also in accordance to our previous research based on PCA and data whitening [4]. We see some countries which are complicated to be predicted and forms separate groups in each case. This group is represented by Bulgaria and Latvia. The country classification serves also as indicator of upcoming crisis to the closest countries.

6 Conclusion

Kohonen SOM learning was used to country self-organization in hexagonal SOM topology with whitened log differentiated macroeconomic data. The best result were obtained for $H = 19$ and 2 dimensional whitening with topological error 0% and minimum possible quadratic penalisation. The resulting SOM maps are in agreement with general expectations. From the crisis prediction point of view there is a group of leading European countries (DE, FR, AT, DK, CY, IE), the other European countries with standard economies (UK, ES, IT, IR, BE, NL, LU, CZ, SK, PL, HU) are in the neighbourhood with slightly different response during crisis. The countries with extreme behaviour during crisis (RS, BG, LV, LT, ME, RO) are placed far from the previous groups. The Kohonen SOM is not too sensitive to dimension of data whitening and therefore, the resulting maps only differ in details but save the country similarity property.

Acknowledgements

Supported by the grants SGS 17/196/OHK4/3T/14 and SGS 17/197/OHK4/3T/14.

References

[1] Directorate General for Economic and Financial Affairs (ECFIN). Statistical annex to european economy. autumn 2016. Technical report, European Commission, 2016.

[2] D. Graupe. *Deep Learning Neural Networks: Design and Case Studies*. 2016.

[3] L. Hamel. Som quality measures: An efficient statistical approach. In *Proceedings of the 11th International Workshop WSOM 2016*, pages 49–59, Houston, 2016. Springer.

[4] R. Hrebik and J. Kukal. Multivarietal data whitening of main trends in economic development. In

Mathematical Methods in Economics, pages 279–284, Plzeň, 2015. University of West Bohemia.

[5] T. Kohonen. *Self-Organizing Maps*. Springer Series in Information Sciences. Springer Berlin Heidelberg, 2012.

[6] E. Oja and S. Kaski. *Kohonen Maps*. Elsevier Science, 1999.

[7] Georg Pölzlbauer. Survey and comparison of quality measures for self-organizing maps.

[8] A. Rettberg, M.C. Zanella, M. Amann, M. Keckeisen, and F.J. Rammig. *Analysis, Architectures and Modelling of Embedded Systems: IESS 2009, Langenargen*. Springer Berlin Heidelberg, 2009.

Investigating the Impact of a Labour Market Segmentation Using a Small DSGE Model with Search and Matching Frictions

Jakub Chalmovianský[1]

Abstract. In this contribution, I present a small scale DSGE model of a closed economy with labour market segmentation, search and matching frictions, and right-to-manage bargaining. Constructed model stems from the work of Lubik [4], however, was redesigned in order to account for different wage setting processes for different groups of workers. I introduce the model with two groups of individuals distinguished, according to the level of their qualification, to low-skilled and high-skilled, similarly as presented by Porter and Vitek in [7]. The aim of this contribution is to examine possibly dissimilar impacts of exogenous shocks to an economy where workers exhibit different levels of qualification. In this sense, I firstly introduce the model with labour market segmentation that splits the aggregate labour market between the following two sectors: one populated by workers with high level of qualification and the other with less qualified workers. This model is then calibrated to fit the characteristics of the US economy, and its dynamic properties are compared with the Lubik's original model.

Keywords: DSGE model, labour market segmentation, right-to-manage bargaining, search and matching frictions.

JEL classification: C32, E24
AMS classification: 91B51, 91B40

1 Introduction

Assessment of labour markets in the DSGE models, for example from the point of their impact for the economy as a whole, importance for monetary policy, or flexibility, has been studied thoroughly in the recent years, hand in hand with the development of the search and matching frictions in these dynamic models. On the other side, labour market segmentation within the DSGE framework is a relatively new approach, even though it may be of great importance for explaining specific aspects of heterogeneous and far from perfectly competitive labour market.

In this contribution I present a novel approach to model the labour market, inspired by the work of Porter and Vitek [7], who divided the labour market into two different segments (sectors). First sector is populated with low-skilled individuals, in the second the high-skilled workers are situated. The advantage of such division occurs when evaluation of the effect of a labour market policy targeted only on a certain sub-population of workers is of interest. This is also the case for the before mentioned research of Porter and Vitek, who studied impacts of introducing a statutory minimum wage in Hong Kong using a DSGE model with labour market segmentation. Segmentation of workers according to the level of their qualification or skills is often used when assessing the effect of minimum wage, however not within the DSGE approach. Another examples, which do not belong to the DSGE literature, can be found in the research of Cahuc et al. [2] or in the monography by Neumark and Wascher [6].

The aim of this contribution is to integrate the labour market segmentation with a small scale DSGE model with search and matching frictions. I build this model by building upon the work of Lubik [4], which at the same time serves as the benchmark. The augmented model will be later used to assess the importance of labour market institutions for the analysed economy and impact of changes in the institutional setting on labour market performance. Nevertheless, in this contribution I present results from the first stages of the model development. Therefore, the aim of this contribution is to examine possibly dissimilar impacts of exogenous shocks to the economy where workers exhibit different levels of qualification. For this purpose, the model presented in next section is calibrated and its simulations are used to evaluate the dynamics of the model in comparison with the benchmark and to assess possible differences in responses to shocks between both labour market sectors.

2 Model

The model presented in this contribution is build upon the work of Lubik, presented in [4]. Therefore, it represents a small scale DSGE model of a closed economy with search and matching frictions on labour market. However, in

[1]Masaryk University, Faculty of Economics and Administration, Lipová 41a, Brno, Czech Republic, chalmovianskyy@mail.muni.cz

order to differentiate types of labour, and to capture aspects of different levels of qualification between individuals, I refined its original labour market sector. I model two sectors of economy, these two sectors differ by the qualification of their labour force (education, working experience, etc). In the first sector low-skilled labour services are supplied, and in the second one the high-skilled labour supply is placed. This approach is inspired by a DSGE model from [7], where similarly defined labour markets are used for studying the impacts of minimum wage on business cycle.

Otherwise, the structure of the model is fairly standard, consisting of households, and firms, without any explicit monetary authority. Households are populated by the before mentioned two types of individuals. Every member of a household is send out to the labour market. There they look for jobs, if unemployed, or supply labour services if employed. Nevertheless, workers and vacancies do not meet instantaneously. In fact, they are matched in a costly search procedure. This search and matching process implies following features:

- even in the equilibrium there can be a non-zero proportion of household members that are unemployed, even though they are willing to work,
- wages are not determined in a competitive market, but negotiated in a bargaining process between firms and workers.

In what follows, I discuss only specific aspects of my model that stem from the double labour market structure. Namely, I briefly present the structure of households and the optimization problems of their various members, and follow with the discussion on wage determination for the two labour market sectors.

2.1 Households

My model economy is populated by a continuum of households indexed by $k \in [0, 1]$. Members of each household are of two kinds. According to their professional qualification: proportion ϕ^L of household members are low-skilled workers (denoted with superscript u) and $(1 - \phi^L)$ are high-skilled workers (denoted as s), where $0 < \phi^L < 1$. Total employment, $n_{k,t} \in [0, 1]$, in the representative household k can be written as follows:

$$n_{k,t} = \phi^L n^u_{k,t} + \left(1 - \phi^L\right) n^s_{k,t} \tag{1}$$

Except for the qualification, all workers are otherwise identical. Furthermore, as in [5], no heterogeneity in asset holdings and consumption of individual workers or households is assumed. Therefore, indices of specific workers/households are dropped in all equations to come. Representative infinitely living household maximizes the following intertemporal utility function:

$$U_t = E_t \sum_{j=t}^{\infty} \beta^{j-t} \left[\frac{(C_j)^{1-\sigma} - 1}{1 - \sigma} - \chi_j D n_j - D^s n^s_j \right] \tag{2}$$

where β is discount factor satisfying the condition $0 < \beta < 1$, parameter σ represents coefficient of relative risk aversion having non-negative values ($\sigma \geq 0$), χ_j is exogenous stochastic process, which may be identified as an exogenous shock in labour market preferences, $D \geq 0$ is parameter of preferences over disutility from work activities, similarly $D^s \geq 0$ is parameter of preferences over disutility from work specific only for high-skilled sector. This concept establishes additional disutility from work for high-skilled employees that may arise from their higher opportunity costs associated with gaining work experience (more years of schooling, increased work effort, etc). Thanks to this construction, my model distinguishes between wages for low-skilled and high-skilled employees, allocating higher wages for the latter. It is worth noting that the proportion of employed individuals in the representative household of a given qualification is fully determined by a matching function. Hence, household cannot control for its labour supply, which may lead to an involuntary unemployment. Finally, C_t denotes composite consumption index with constant elasticity of substitution between different consumed kinds of goods, i.e.:

$$C_t = \left[(1 - \omega_u)^{\frac{1}{\eta_c}} \left(C^s_t\right)^{\frac{\eta_c - 1}{\eta_c}} + (\omega_u)^{\frac{1}{\eta_c}} \left(C^u_t\right)^{\frac{\eta_c - 1}{\eta_c}} \right]^{\frac{\eta_c}{\eta_c - 1}} \tag{3}$$

with ω_u denoting the fraction of consumed goods produced by low-skilled workers and η_c as the elasticity of substitution between the two kinds of goods. This implies the following demand functions for each type of good:

$$C^s_t = (1 - \omega_u) \left[\frac{P^s_t}{P_t} \right]^{-\eta_c} C_t \tag{4}$$

$$C^u_t = \omega_u \left[\frac{P^u_t}{P_t} \right]^{-\eta_c} C_t \tag{5}$$

where P_t^s and P_t^u are aggregate price indices for the high-skilled and low-skilled production respectively, and P_t is the composite price index, defined as their weighted average, i.e.:

$$P_t = \left[(1-\omega_u) \left(P_t^s\right)^{1-\eta_c} + \omega_u \left(P_t^u\right)^{1-\eta_c} \right]^{\frac{1}{1-\eta_c}} \tag{6}$$

Finally, the representative household's budget constraint is as follows:

$$C_t + T_t = w_t^s n_t^s + w_t^u n_t^u + (1 - n_t^u)\, b + (1 - n_t^s)\, b + \Pi_t \tag{7}$$

where T_t are lump-sum taxes used to finance the unemployment benefits b, Π_t are profits that representative household receives as the owner of the firms, and w_t^l is the real wage that each worker of given qualification receives for supplying her labour services.

Employment status of each worker (member of household) is fully determined by the outcome of the search and matching process. Thus, there is no explicit way for the household to choose the level of labour supplied. The only problem of households is to select the amount of consumed goods of each kind. Taking into account the fact that there is no intertemporal aspect, the resulting first order condition is of the following form:

$$(C_t)^{-\sigma} = \lambda_t \tag{8}$$

where λ_t denotes Lagrange multiplier for the household's budget constraint.

2.2 Wage bargaining

Labour market is specified, as mentioned earlier, using the search and matching process with households supplying labour services to firms on a frictional market. Details of this process can be found in Lubik's contribution [4]. The only difference is in the use of separate labour market for each of the two groups of workers. On the other hand, the wage bargaining process reveals some differences for each kind of labour force.

Recall crucial assumption of this model: labour force can be divided into two separate groups based on the qualification level of its members. Individuals in each of these two populations are entering qualification-specific labour market and participate in production of specific goods. Except for this difference all workers are otherwise identical, their qualification casts the only contrast. If this assumption holds, wage of the qualified worker results from the bilateral bargaining process between firms and representatives of the labour force. Both participants in the negotiations try to set the wage rate, w_t^s, in order to maximize the total surplus generated by the worker-firm linkage. The total surplus is split to maximize weighted average of the individual surpluses of firm and worker. The bargaining function, S^s, used in this model is of the following form

$$S_t^s \equiv \left(\frac{1}{\lambda_t} \frac{\partial \mathcal{W}_t^s(n_t^s)}{\partial n_t^s} \right)^{\eta^s} \left(\frac{\partial \mathcal{I}_t^s(n_t^s)}{\partial n_t^s} \right)^{1-\eta^s} \tag{9}$$

where η^s represents the workers' weight in negotiations, $\frac{\partial \mathcal{W}_t^s(n_t^s)}{\partial n_t^s}$ is the marginal value worker contributes to the household's welfare, and from the firm's perspective $\frac{\partial \mathcal{I}_t^s(n_t^s)}{\partial n_t^s}$ is marginal value worker contributes to the firm's welfare. The latter determined by the firm's first order condition with respect to the requested number of high-skilled workers.

Marginal contribution of the high-skilled worker for the household's welfare, $\frac{\partial \mathcal{W}_t^s(n_t^s)}{\partial n_t^s}$, can be obtained making use of the options the worker has. If employed, this contribution equals the wage worker earns, w_t^s. However, at the same time she suffers from the disutility from work represented by $\chi_t D n_t$, as well as the additional disutility specific for the high-skilled work, $D^s n_t^s$. At last but not least, she loses the unemployment benefits, b. Additionally, the worker's marginal contribution to the welfare of her household depends also on its expected value in the next period. Note that the real payments are again valued at the marginal utility λ_t. The marginal value of a high-skilled worker for the household is:

$$\frac{\partial \mathcal{W}_t^s(n_t^s)}{\partial n_t^s} = \lambda_t w_t^s - \lambda_t b - D\chi_t - D^s + \beta E_t \frac{\partial \mathcal{W}_{t+1}^s(n_{t+1}^s)}{\partial n_{t+1}^s} \frac{\partial n_{t+1}^s}{\partial n_t^s} \tag{10}$$

Assuming the following evolution of employment in time: $n_t^l = (1-\rho^l)\left[n_{t-1}^l + v_{t-1}^l q(\theta_{t-1}^l)\right]$, the expression $\frac{\partial n_{t+1}^s}{\partial n_t^s}$ can be replaced with: $(1-\rho^s)\left[1 - \xi^s \theta_t^s q(\theta_t^s)\right]$. Subsequently, the derivation of the resulting function, S_t^s, with respect to the only variable that is subject to bargaining (i.e. the wage rate of the high-skilled workers,

w_t^s), applying the expressions for the worker's marginal contributions to household's and firm's value respectively, results in the following optimality condition of the bargained wage:

$$w_t^s = \eta^s \left[\alpha \frac{y_t^s}{n_t^s} \frac{\varepsilon_t}{1+\varepsilon_t} p_t^s + \kappa^s \left(v_t^s\right)^{\psi^s - 1} \theta_t^s \right] + (1-\eta^s) \left[b + \left((1-\phi^L) D \chi_t + D^s \right) \left(C_t^s\right)^{\sigma} \right] \tag{11}$$

The bargained wage is simply a weighted average of the payments accruing to workers and firms. Among other it includes mutual compensation of the job related costs, e.g. hiring costs, $\kappa^s \left(v_t^s\right)^{\psi^s - 1}$, or costs in the form of utility loss of employed, which have two components: the loss of utility from leisure, $\left((1-\phi^L) D \chi_t + D^s \right) \left(C_t^s\right)^{\sigma}$, and the unpaid unemployment benefits, b.

In the case of the low-skilled workers, denoted by the superscript u, I assume slightly different scenario. The bargained wage is specified by analogous bargaining process as in previous case, except with different parameters that are typical for the low-skilled sector (particularly with different term for the workers' loss of utility from leisure). The resulting bargained wage is then defined by the following optimality condition:

$$w_t^u = \eta^u \left[\alpha \frac{y_t^u}{n_t^u} \frac{\varepsilon_t}{1+\varepsilon_t} p_t^u + \kappa^u \left(v_t^u\right)^{\psi^u - 1} \theta_t^u \right] + (1-\eta^u) \left[b + \phi^L D \chi_t \left(C_t^u\right)^{\sigma} \right] \tag{12}$$

2.3 Closing the model

In order to close the model following social resource constraints, specific for each labour market segment, need to be defined:

$$Y_t^s = C_t^s + \frac{\kappa^s}{\psi^s} \left(v_t^s\right)^{\psi^s} \tag{13}$$

$$Y_t^u = C_t^u + \frac{\kappa^u}{\psi^u} \left(v_t^u\right)^{\psi^u} \tag{14}$$

Finally, I assume the logarithms of four exogenous stochastic processes, namely the technology shock, A_t, the labour shock, χ_t, the matching efficiency shock, μ_t, and the time-varying demand elasticity, ϵ_t, to follow independent $AR(1)$ processes with coefficients ρ_i, $i \in (A, \chi, \mu, \varepsilon)$ and innovations $\epsilon_t^i \sim \mathcal{N}(0, \sigma_i^2)$.

3 Calibration and simulation results

The aim of this contribution is to examine possibly dissimilar impacts of exogenous shocks to the economy where workers exhibit different levels of qualification in comparison with the model where no such diversification is introduced. The model presented in this work is developed on the basis of Lubik's model [4]. Therefore, this will serve as a natural benchmark for comparison. For keeping the starting line of both models as similar as possible, many of the model parameters are calibrated using the same values, mostly proceeding from those used by Lubik. However, following steady-state values for unemployment and vacancy rates, and the proportion of low-skilled workers in households are specified with respect to actual characteristics of US economy, as the average values in data over the observed period[2]: $\bar{u} = 0.0639$, $\bar{u}_u = 0.1318$, $\bar{u}_s = 0.0639$, $\bar{v} = 0.0256$, $\bar{v}_u = 0.0748$, $\bar{v}_s = 0.0174$, $\phi_L = 0.245$. Table 1 summarizes values of all calibrated parameters. All computations have been performed using Dynare toolbox (version 4.4.3) for MATLAB developed by Adjemian et al.[1].

In Figure 1 I present the impulse response functions (IRFs) for the following three variables: vacancies, v_t, wages, w_t, and number of successful matches, m_t; in reaction to three shocks (technology shock, A_t, labour shock, χ_t, and matching efficiency shock, μ_t) for two different models. First model depicted by black line is the Lubik's original version and, blue together with red line refer to the high-skilled and low-skilled sector in my augmented model with segmented labour market respectively. Since I do not have any estimates of standard deviations for the stochastic processes, I set the size of these shocks to positive one. Thus one must take into account the incomparability of IRFs amplitudes and the speed of return to a equilibrium between the original and my model. Nevertheless, the direction of the effect for each shock-variable pair can be discussed. As for the technology shock, my augmented model shows hump-shaped reactions of the variables of interest with increasing positive gap in first quarters after the shock hits modelled economy and gradual closing of this gap later on. In comparison with Lubik's model expected effect of the shock comes with a delay of 5 to 10 quarters. For the labour shock the direction of IRF in the augmented model is fairly the same as in the benchmark model for all presented variables, only for the high-skilled sector we can see more wavy reaction, the variables of interest are

[2]Values are based on the US data for period 2001Q1 - 2016Q3 with following sources: Harmonized Unemployment Rate (15 years and over, SA) - OECD database, Unemployment Rate - Less than a High School Diploma (16 years and over, NSA) - BLS, Job Openings data (SA) - BLS, Incidence of Low Pay - OECD.

Parameter	Description	Value		
		Lubik	Model - 2 labour markets	
β	Discount factor	0.99	0.99	
α	Labour elasticity	0.67	0.67	
σ	Relative risk aversion	1	1	
ξ	Match elasticity	0.7	$\xi^u = 0.7$	$\xi^s = 0.7$
ρ	Separation rate	0.1	$\rho^u = 0.1$	$\rho^s = 0.1$
η	Bargaining power of workers	0.5	$\eta^u = 0.5$	$\eta^s = 0.5$
b	Unemployment benefits	0.4	0.4	
ψ	Elasticity of vacancy creation costs	1	$\psi^u = 1$	$\psi^s = 1$
κ	Scaling factor on vacancy creation	0.05	$\kappa^u = 0.05$	$\kappa^s = 0.05$
ω_u	Fraction of consumed goods produced in u sector	N/A	0.245	
D	Scaling factor on disutility from work	N/A	0.35	
D^s	Scaling factor on disutility from work in s sector	N/A	0.65	
$\rho_{\{A,\chi,\mu,\varepsilon\}}$	AR coefficient of shocks	0.9	0.9	
$\sigma_{\{A,\chi,\mu,\varepsilon\}}$	Standard deviation of shocks	1	1	

Table 1 Calibrated parameters

oscillating around the equilibrium. In fact, this behaviour of high-skilled sector variables is present in case of shock to matching efficiency as well, whereas low-skilled vacancies, wages and matches converge to the equilibrium in much more straightforward way. The possible explanation for this may be a stronger spillover effect from the low-skilled to high-skilled sector. Secondly, the direction of the IRFs for μ_t is different for the augmented model with comparison to the benchmark. At first, higher matching efficiency leads to a significant decline in number of vacancies for both sectors, as they are quickly occupied on more flexible labour market, and only after few quarters the number of vacancies rise above the equilibrium level, even though only slightly in the case of unskilled sector. This is in line with the reaction of wages, as they decline well below the equilibrium at the first place, and then gradually rise up in a similar manner, even outperforming the equilibrium in the high-skilled sector again. From the perspective of number of matches, it seems that the decline in wages at the very beginning balances the initial decline in number of vacancies and hence no change in the number of successfully matched workers and vacancies is taking place. However, a positive gap in m_t opens right after that.

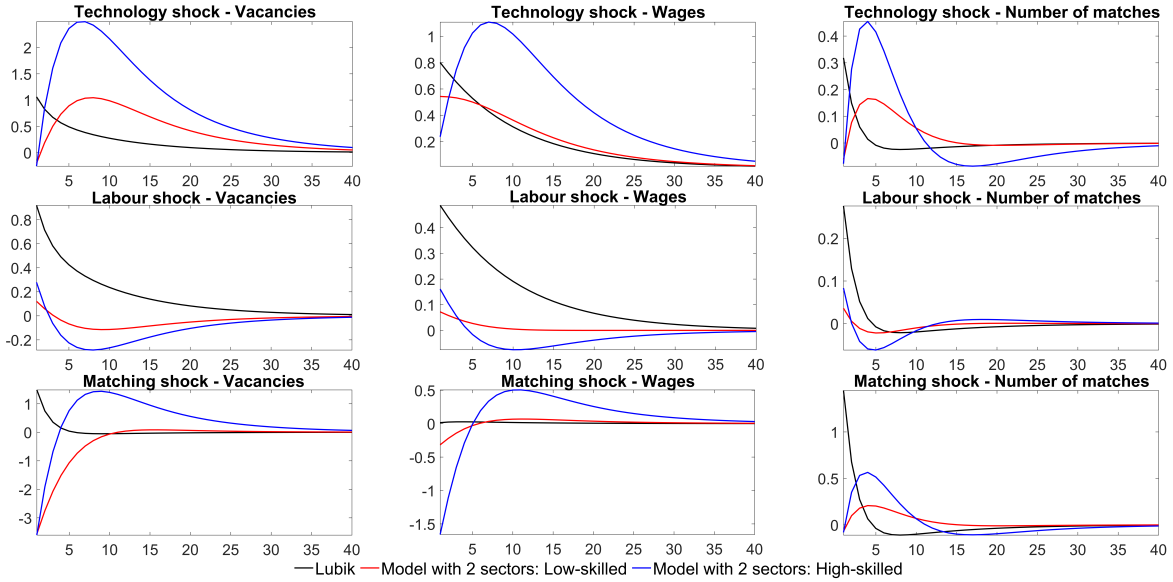


Figure 1 Impulse response functions for selected model variables

Figure 1 allows us to evaluate the difference between skilled versus unskilled sector as well. First findings of this comparison are the already mentioned fluctuations in the convergence process of variables representing the skilled sector, that may indicate presence of a spillover effect from unskilled to skilled sector. Beside these

fluctuations around zero, size of the effects is the second typical difference. In the skilled sector, response to a shock is of greater intensity and it takes more time for the effect to disappear. If we define labour market flexibility as the ability of the economy to absorb exogenous shocks, then these results indicate slightly less flexible labour market for the skilled workers. Together with the size of this market, which is calibrated to constitute approximately three quarters of the whole economy, its flexibility significantly affects the economy as a whole.

4 Conclusion

In this contribution I present a small-scale DSGE model with search and matching frictions on the labour market, which is defined using a novel approach. The labour market is divided into two separate sectors, defined specifically for high-skilled and low-skilled individuals. The model is calibrated in the similar way as the one presented by Lubik [4], which I build upon, in order to see the possibly different behaviour of selected labour market variables in response to exogenous shocks. Although in this contribution only the results from a simulation of the model are presented, some differences from the benchmark can already be seen. However, they are still in a reasonable fashion. The main finding from the comparison between low- and high-skilled sector is the more sensitive reaction of the latter leading to a slightly slower process of absorption of the shocks. However, as this is still only a preliminary work, there are many issues for thge future research. Some of the main questions still to be addressed are: the explanation for the wavy reaction of some high-skilled variables to shocks, or proper estimation of the model using the real data of US economy. The labour market segmentation may be helpful when specific aspects of labour market acting differently on various heterogeneous agents of this market are of interest, or when effects of labour market policies affecting only a small sub-population of workers are to be evaluated, e.g. the effect of change in minimum wage.

Acknowledgements

This work was supported by funding of specific research at Faculty of Economics and Administration, Masaryk University, project MUNI/A/0987/2016. This support is gratefully acknowledged.

References

[1] Adjemian, S., Bastani, H., Juillard, M., Karam, F., Mihoubi, F., Perendia, G., Pfeifer, J., Ratto, M., and Villemot, S.: *Dynare: Reference Manual, Version 4*. Dynare Working Papers, 1, Cepremap.

[2] Cahuc, P., et al.: The consequences of the minimum wage when other wages are bargained over. *European Economic Review* **45** (2001), 337–352.

[3] Krause, U. M., and Lubik, T.: The (ir)relevance of real wage rigidity in the New Keynesian model with search frictions. *Journal of Monetary Economics* **54** (2007), 706–727.

[4] Lubik, T.: Estimating a Search and Matching Model of the Aggregate Labor Market. *Economic Quarterly* **95** (2009), 101–120.

[5] Merz, M.: Search in the Labor Market and the Real Business Cycle. *Journal of Monetary Economics* **36** (1995), 269–300.

[6] Neumark, D., and Wascher, W. T.: *Minimum wages*. MIT Press, Cambridge, Mass., 2008.

[7] Porter, N., and Vitek, F.: *The Impact of Introducing a Minimum Wage on Business Cycle Volatility: A Structural Analysis for Hong Kong SAR*. Working Paper WP/08/285, International Monetary Fund, 2008.

[8] Thomas, C.: Search Frictions, Real Rigidities, and Inflation Dynamics. *Journal of Money, Credit and Banking* **43** (2011), 1131–1164.

Dynamic Modeling Economic Equilibrium with Maple

Zuzana Chvátalová[1], Jiří Hřebíček[2]

Abstract. We present a modern approach to solution of mathematical models described by standard differential equations with delayed argument with using the Maple™ system. Currently, the library package with resources for differential equations with delay is built in the Maple system. We (briefly) discuss this advanced library package as another sub-goal of this paper.
We present a background for formulation of interactive numerical simulation – modeling of price speed of changes in time as a continuous dynamic model of the partial economic equilibrium (of the demand and supply) on the market with perfect competition. Naturally, we assume that in the market economy change of variables is dependent on the variables values, which was achieved in the past. Standard differential equations with delay can describe the time development of real phenomena more precisely. In the case study, we designed an interactive Maplet that visualizes computational simulation.

Keywords: differential equations with delay, economic equilibrium, Maple.

JEL Classification: C02, D58
AMS Classification: 34N05

1 Introduction

At present, economical theories as well as empirical economical research are both being visited by quantitative methods, which are working to resolve the issue of the economic reality, more and more efficiently and with higher acclaim. This is occurring whilst various areas of mathematics and statistics are being utilized (modelling, dynamic programming, optimization, econometrics, and others), as has been confirmed by, for instance [4],[7],[9],[21]. Information and communication technologies (ICT), which are being applied more and more, have been contributing to this.

"*Economical data is often generated by processes that have a dynamic character."* [1]. Historically as well as nowadays, the issue of dynamic systems has become the aim of many solutions of practical as well as theoretical tasks, e.g., in [6],[23]. Dynamic models (systems) have been respecting the development of their state in time. The state of a system ought to be understood as a vector of state variables and the system's behavior (state change) should be taken as their formalized mathematical description. Including the time factor into the model is pre-determined by the option of following the speed of the model's state change while describing it through differential equations, or rather their systems (more about applications for example in [18]).

Let us say that – as has been noted in detail in [11] – natural laws and also empirical research of real phenomena have been proving that changes of independent variables (explaining economical phenomena) are often implemented in sequence of "temporal intervals". The dependent (explained) variable then changes based on the development of past values. The reason for this is most often the slow response of the economic actors (all in all, they are unable to adapt quickly). Therefore, during the construction of a dynamic model, it is reasonable to consider past values (the so-called delay) of the variables as well.

This contribution deals with the dynamic joint modelling of the partial market balance of demand *D* and supply *S*, i.e., on the partial single-commodity market of goods and services with perfect competition (according to Holman, this is the market with perfect information of the buyers, zero costs for the change of supplier, a homogeneous product). We respect the temporal development of the changes in quantity and price. This is why we used, for the purpose of project resolution, the theory of functional differential equations. The modelling is done in a modern way through differential equations with a delayed argument (DDE) with the use of the Maple system as a part of the Czech Science Foundation project. Name of the Project:" Development of new methods of solving dynamic models of corporate processes management" GA16-03796S.

[1] Brno University of Technology, Faculty of Business and Management, Department of Informatics, Kolejni 4, 612 00 Brno, Czech Republic, e-mail: chvatalova@fbm.vutbr.cz
[2] Masaryk University, Institute of Biostatistics and Analyses, Kamenice 126/3, 625 00 Brno, Czech Republic, e-mail: hrebicek@iba.muni.cz

Including the delay states more precisely the solution of the differential equations. Approximately since the half of the 20th century, the development of DDE began, being meant to be applied in various fields, e.g., [13],[14],[17],[22], etc.

Numerical simulation of the case study is implemented through the Maple system which boasts a number of advantages. Among others, this system performs not only numerical, but also symbolic calculations; it features an elaborate **dsolve library** designated for the resolution from the sphere of differential equations. Since its 2015 version, Maple contains, in the **dsolve[numeric],** embedded and pre-defined means of resolving real problems using DDE.

2 Maple: Dsolve Library - Introduction

The Maple system has been developed by the company Maplesoft [16] since 1990. Maplesoft's core technologies include the advanced computation engine in Maple environment and physical modelling techniques in MapleSim environment. Combined together, these technologies enable the creation of cutting-edge tools for design, modelling, and high-performance simulation. There is implemented a real-valued delay differential initial value problem in Maple. Maple finds a numerical solution for the general delay initial value problems. It is able to detect the presence of delays automatically, but it requires additional information for the case of variable delays in the procedure **dsolve/numeric/delay** [15]. Calling sequence for numerical solution of differential equations with delay in Maple is command:

`>dsolve(delaysys,numeric,vars,options),` where `dsolve` command parameters means: `delaysys` - set or list; ordinary differential or differential-algebraic equation(s) with delays; `numeric` - name; instruct dsolve to find a numerical solution; `vars` - (optional) dependent variable or a set or list of dependent variables for parameter `delaysys`;`options` - (optional) equations of the form `keyword=value`.

Description: The `dsolve` command with the parameter `numeric` or `type=numeric` option and a real-valued delay differential initial value problem finds a numerical solution for the delay of initial value problem.

Options: If the optional equation `method=numericmethod` is provided (where paremeter numericmethod is one of rkf45[3], ck45[4], or rosenbrock[5]), dsolve uses that method to obtain the numerical solution. There are used numerical methods: rkf45 - the Fehlberg fourth-fifth order Runge-Kutta [5]; ck45 - a variable order Runge-Kutta Method for initial value problems with rapidly varying right-hand sides [2] or rosenbrock - the implicit Rosenbrock third-fourth order Runge-Kutta [20] to obtain the numerical solution.

The `dsolve` command can detect the presence of delays automatically, but it requires additional information for the case of variable delays (see parameter `delaymax` later). In addition to the options `delaymax` and `delaypts` described in command `dsolve[numeric]`[6] and `dsolve[Error_Control]`[7], delay differential systems also have the following options: `delaymax=numeric`; `delaypts=positive integer`, where the option `delaymax=numeric` is required for problems with a variable delay, and is required to specify the maximum delay time for all delays in the system and the option `delaypts=positive integer` specifies the maximum point storage to allocate when storing delay data, and must be at least 6, and at that setting will only be able to retain an accurate solution history for the last integration step taken. By default, this value is conservatively set to 10000. Note that if a small value is used, the computation will use less memory, but the results may be inaccurate. A warning is produced if this value is specified below 100.

Numeric solution of delay: The delay computation implemented for the Maple numeric solvers utilizes the natural interpolant built into the solvers, storing solution values back to `delaymax` as the computation proceeds. On initialization, the dependent variable values are assumed to be constants set to the initial values for all times $t < t_0$, where t_0 is initial value of independent value, and once the delay time for a delay term exceeds $t - t_0$, the interpolant comes into play. Derivatives of the dependent variable values are assumed to be zero for all times $t < t_0$. In some cases, such as delays for derivatives, use of the initial value or zero is not desirable, and in these cases a constant value for $t < t_0$ can be specified as the second argument to the delay function. For example, $y(t - 1)$ specifies that for $t < 1$ the initial value of y is used, while $y(t - 1, 0)$ specifies that for $t < 1$ the value 0 be used. Similarly, $\frac{d}{dt}(y)(t - 1,0)$ specifies that for $t < 1$ the value 0 be used, while $\frac{d}{dt}(y)(t - 1,1)$ specifies that for $t < 1$ the value 1 be used. In the event that insufficient storage is allocated for a problem (in other words, if `delaypts`

[3] http://www.maplesoft.com/support/help/Maple/view.aspx?path=dsolve/rkf45
[4] http://www.maplesoft.com/support/help/Maple/view.aspx?path=dsolve/ck45
[5] http://www.maplesoft.com/support/help/Maple/view.aspx?path=dsolve/rosenbrock
[6] http://www.maplesoft.com/support/help/Maple/view.aspx?path=dsolve/numeric
[7] http://www.maplesoft.com/support/help/Maple/view.aspx?path=dsolve/numeric/Error_Control

was set too small), the storage downgrades. Initially, it reduces to storing only integration step endpoints using a linear interpolant. Then if further downgrading is required, the storing reduces to every second endpoint, and consequently to every fourth, every eighth and so on. Note that solution accuracy is significantly degraded when downgrading is in use, so it is best to make `delaypts` sufficiently large so that downgrading is unnecessary.

3 Market Description

The approaches to the problems of market balance have been developing through time. Generally, the known ideas of Adam Smith, e.g., the idea that the free market itself seeks to the equilibrium via the effect of market mechanisms, have been followed upon by important economists of the 19th and 20th century like Alfred Marshall, who deals with the idea of supply and demand, marginal utility and production costs as a whole; or Léon Walras, who moves from partial balance to general balance. The most important economists dealing with, among others, general equilibrium includes John Richardo Hicks (awarded the Nobel Prize in 1972). The founder of the ordinalistic approach to limit use is Vilfredo Pareto. At present, hand in hand with the utilization of ICT means, a whole range of scientists are researching balance, e.g., [1].

We will construct a dynamic continuous model of partial economic equilibrium of the market demand *D* and the market supply *S* at a partial single-commodity goods and services market with perfect competition (according to Holman, this is a market where the buyers are perfectly informed, there are zero costs of changing the supplier, there is a homogeneous product) [10], more also in [24].

Let us consider the market demand and the market supply as a sum of all the individual demands and a sum of all the individual supplies on this market. Then let us express the single-factor model of demand as a function $D: R^+ \to R^+$, abiding by the law of decreasing demand (ceteris paribus) and a single-factor model of supply as a function $S: R^+ \to R^{0,+}$ (companies offer a commodity for the price which is at least the level of the costs), abiding by the law of increasing supply (ceteris paribus). They are models with the endogenous variables of the P price and Q quantity, which depend on time t. We will suppose that both of the functions are continuously differentiable in their domain.

Remark: We will, in the description, follow the diction of natural language: for our models, the price P is the independent variable and the quantity Q is the dependent variable: $Q_D = D(P)$ a $Q_S = S(P)$, Q_D is the demanded quantity, Q_S is the supplied quantity. In the visualization: P – horizontal axis, $Q_{D,S}$ – vertical axis. It holds that $\frac{dD}{dP} < 0$ and $\frac{dS}{dP} > 0$.

The market equilibrium is a market state: there is a commodity P price, on the level of which companies offer exactly the amount of the given commodity which the buyers are willing to demand for this price, i.e., $Q_D = Q_S$. Then we can speak of the equilibrium price and equilibrium quantity, we mark this as $[P^E, Q^E]$. The only stable state of a market is the market equilibrium. There is a single equilibrium price (market clearing), from which the actual price can – via the effects of market forces – deviate. A stable dynamic system is such a system which, from a relevant "disequilibrium", naturally converges to equilibrium. Then the economic subjects react to the equilibrium which is a temporary state, because it generates a price change moving towards equilibrium. The stable market then returns to equilibrium according to the theorem of the cobweb. It is known (in accordance for example with [19]), that the trajectory of this process converges towards balance, if, in the absolute value, the slope of the supply curve is smaller than, in the absolute value, slope of the demand curve, i.e. (1):

$$\left|S'(P)\right| - \left|D'(P)\right| < 0 \tag{1}$$

The market disequilibrium is characterized by two states:

- $D(P) - S(P) > 0$: overshooting of the demanded quantity versus the supplied one, $Q_D > Q_S$, the current commodity price P is lower than the equilibrium price, $P < P^E$. Companies respond by increasing the price and increasing production.
- $D(P) - S(P) < 0$: overshooting of the supplied quantity versus the demanded one, $Q_S > Q_D$, the current commodity price P is higher than the equilibrium price, $P > P^E$. Companies respond by decreasing the price and decreasing production.

4 Stages of Modeling a Continuous Dynamic Market Equilibrium Model

Firstly, we consider models of demand and supply as *the stationary fixed linear functions* (2):

$$Q_D = D(P) = a \cdot P + b, \qquad Q_S = S(P) = c \cdot P + d \tag{2}$$

where a, b, c, d are real parameters of the model $a < 0$ (the law of decreasing demand), $c > 0$ (the law of increasing supply), and $b > d$.

Such models can be constructed using empirically observed data using various mathematical techniques (numerical, statistical, optimization methods).

In the market equilibrium: $Q^E = Q_D = Q_S$ at a certain price P^E. Then from (2) we can deduce that the market equilibrium point is equal to $[P^E, Q^E] = [\frac{b-d}{c-a}, \frac{cb-ad}{c-a}]$. Regarding the equilibrium model convergence, the condition $|c - a| < 0$ must be met.

Secondly, to create *a dynamic model*, we assume that the price $P(t)$ changes continuously depending on the real time t; the speed of price change is directly proportional to the difference between the currently demanded and the currently supplied quantity, i.e., $\frac{dP}{dt} = g \cdot (Q_D - Q_S)$, where $g \in R^+$ is the proportionality constant reflecting the market sensitivity. Thus, a differential equation (3) can be written as follows:

$$\frac{dP}{dt} = g \cdot (a \cdot P(t) + b - c \cdot P(t) - d) = g \cdot ((a - c) \cdot P(t) + (b - d)) \tag{3}$$

Remark: Using the equilibrium point (1) can be converted: $\frac{dP}{dt} = g \cdot ((P(t) + P^E) \cdot (a - c))$.

As a solution of this differential equation in Maple, we receive: $P(t) = P^E + e^{g(a-c)t}\_C1$, where $\_C1$ is a real constant. After choosing the initial condition, $P(0) = P^0$ is $\_C1 = P^0 - P^E$. The solution of this differential equation (3) has this form: $P(t) = P^E + e^{g(a-c)t}(P^0 - P^E)$.

Thirdly, for the subsequent formalization of *a continuous dynamic model*, we begin at a case of discreetly varying time. We can derive a dynamic equilibrium model (differential equation) from the following demand and supply functions in the form (4):

$$Q_D^t = D(P) = a \cdot P^t + b, \qquad Q_S^t = S(P) = c \cdot P^{t-1} + d \tag{4}$$

with the parameters a, b, c, d (the same as in (2)), Q_D^t, Q_S^t means monitored supply and demand in time points $t = 0, 1, 2, 3, \ldots$. Supplied quantity Q_S^t in time t is governed by the price P^{t-1} in time $t - 1$. I.e., companies set the price according to the previous period.

Therefore, fourthly, the discrete time dynamic model (4) is a motivation to introduce the time delay τ (that is the time needed to supply change relative to the price development in the past). Then we get a *differential equation with a delayed argument* (5):

$$\frac{dP}{dt} = g \cdot (a \cdot P(t) + b - c \cdot P(t - \tau) - d)) = (a \cdot P(t) - c \cdot P(t - \tau) + (b - d)) \tag{5}$$

where a, b, c, d, g are model *parameters*, $\tau > 0$ is the time delay and t is time as a continuous variable (all with the above properties).

To solve the model (5), it is necessary to determine its behavior at two time intervals $I = <-\tau, 0)$ and $J = <0, t_0>$, where t_0 is some future time value. For the supply in the time interval I, we have to define the so-called "historical function" $P_H(t - \tau)$. Supply in the time interval J is governed by a "non-historical" function $P(t - \tau)$. In the model (5), they can be applied using the indicator function: $\chi(t) = 0$ for $t \in I$ and $\chi(t) = 1$ for $t \in J$. The initial condition introducing $P_H(0) = P(0)$ ensures that the condition of the continuous continuity of the historical and non-historical function at point $t = 0$ is fulfilled.

We can find evidence of the existence, uniqueness of solution and convergence, respectively, of the stability of the models of such types in literature, for example in [8],[12].

The formalized (explicit) form of the expected solution cannot be directly expressed. We can use due to solution through the iterative process, for example, the step method or the successive approximations method. By using the appropriate executive software, in our case the Maple system, we can achieve a quick solution, preferably by visualizing the whole situation.

5 Case Study

In this paper, we are dealing with a partial market with a single type of goods in perfect competition. Such a market is rather theoretical for the microeconomic interpretation of market functioning (i.e., for example, the grain, corn, cars markets, etc.) The model (5) is implemented in the hypothetical wheat market in the Czech Republic. Due to the difficult availability of direct data, we had to estimate and modify it very much. Also due to

the economic crisis in the beginning of the monitored years, it is necessary to consider the distortion of the primary converging model (2). We see it as an illustrative application of the above. Therefore, we have compiled a common Maplet, in order to update the data operationally (Fig. 1). The linearity of the input models, the delay implementation, and the required initial condition enforce 5 + 1 + 1 = 7 parameters in the Maplet construction:

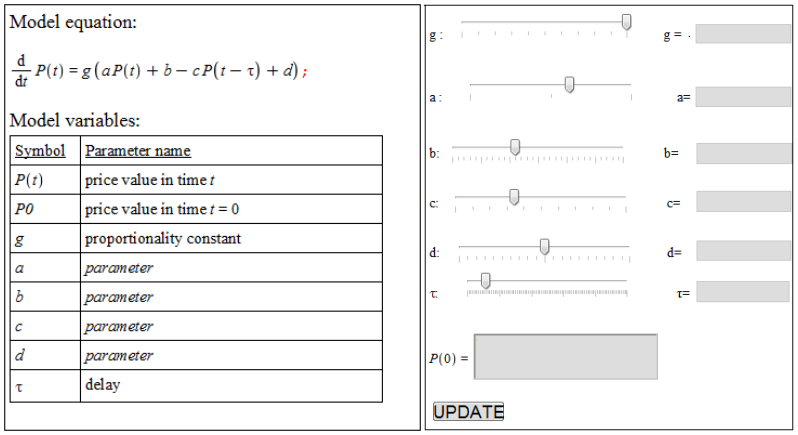


Figure 1 The Maplet - interactive equilibrium model (Source: Authors' own work in Maple)

We have shown how the Maple system is able to handle advanced issues with the implementation of differential equations with a delay. We have used the basic data (in the years 2009 to 2015 - we have chosen it, due to the availability of data and the delay determining) from the available Czech Statistical Office database [3] and by direct expert assessment (which does not want to be named) to determine the price equilibrium for establishing the balance price for consumption wheat for one person per year. Our reflections were based on the main idea of price of consumed food wheat per 1 kg per person in that year ([3], especially and in particular, we used website with the time series of agriculture surveys, surveys of selected products, harvest of wheat, consumer price research, etc., and expert assessments.) Unfortunately, to gain data that would speak exactly is very difficult, so we have used a range of modifications. We focused on finding the price equilibrium [CZK / kg]. (According to [18], it is possible to extend the ideas by solving a system of two equations with two unknowns and also to determine the equilibrium quantity.)Thus, the model (2) can have a somewhat non-typical form, but it ensures the basic features – decreasing demand and increasing supply, while the cobweb convergence holds. To obtain the parameters for model (2) we have used the regression analysis tools in Maple. Subsequently, we have constructed a model (5) with a constant delay:

$$\frac{d}{dt}P(t) = g \cdot (40{,}6093828 - 2{,}3896306 \cdot P(t) - 0{,}3416628 \cdot P(t-\tau)$$

We understand the proportionality ratio g to be an unknown value. We have tested the measure of its contribution to the solution. We have obtained the graphical solution. Fig. 2 shows selected solutions for different options of g, $\tau = 1$ and the initial condition $P(0) = 11{,}03$. It is also the historical function $P_H(t-1) = 11{,}03$ (constant delay).

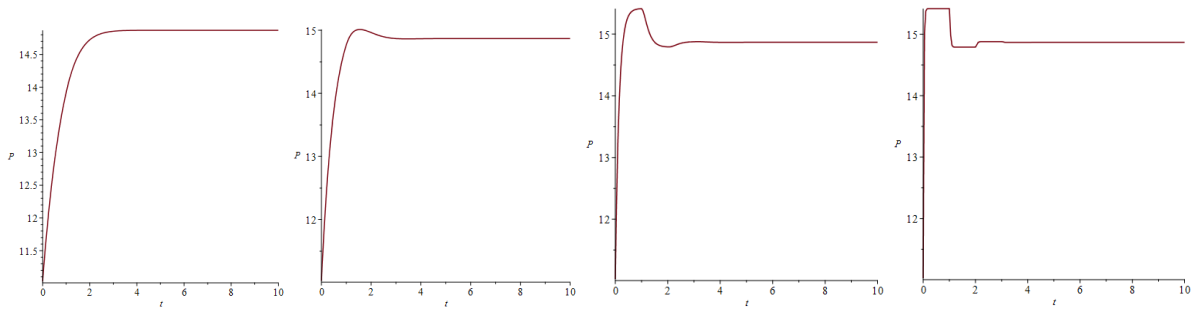


Figure 2 Solution - from the left: $g = 0{,}45$; $g = 1{,}1$; $g = 3{,}0$; $g = 20{,}0$ (Source: Authors' own work in Maple)

We can see the sensitivity of the model to achieve an equilibrium rate price for various values of g. The other choice for τ and the historical function has an additional effect on the outcome. In this paper, we selected $\tau = 1$ and$P_H(t-1) = 11{,}03$. We were considering $g \in (0{,}20)$. For our hypothetical market, we were observing that for: $g < 0{,}45 \rightarrow$ the solution tends to the equilibrium from the time $t \cong 2$ as an increasing concave function graph; then, increasing the g value approximately to 1,1 gives slight oscillations around the equilibrium in the later time (from $t \cong 3$); further increasing the g value approximately to 3 gives more distinct oscillations around the equilibrium; g values even higher than 3 cause sharp oscillations around the equilibrium in the later time ($t > 3{,}5$).

6 Conclusion

The paper presents the possibilities of the Maple system to visualize a dynamic model of price equilibrium (demand, supply) in a partial market with perfect competition and one commodity. We used a differential equation with a delay to describe the iterative process of a simulated convergent cobweb model. Due to the difficulties (generally) to obtain data in economic practice or to the possibility of specifying changes of the input parameters operationally, we have constructed an interactive Maplet. We were testing this Maplet for a hypothetical wheat market. We have found that the iteration process is very well presentable in this software environment. We have verified our model always at a certain level of delay and expression of the historical function (i.e. ceteris paribus) by gradually changing the unknown value of the proportionality constant and we have monitored the effect of these changes on the model's convergence to the equilibrium price. In this paper, we selected $\tau = 1$ and $P_H(t-1) = 11{,}03$. If we add, from the model (2), the second equation in which we express the quantity Q in relation to time t, we will obtain (as an added bonus) – by solving a system of two equations with two unknowns – as an iterative process in Maple, the quantity equilibrium.

Acknowledgements

This paper was supported by the Czech Science Foundation. Name of the Project:" Development of new methods of solving dynamic models of corporate processes management" GA16-03796S.

References

[1] Anton, J. J., and Biglaiser, G. Quality, upgrades and equilibrium in a dynamic monopoly market. *Journal of Economic Theory*, **148**/3(2013), 1179-1212.
[2] Cash, J.R.; Karp, A.H.: A Variable Order Runge-Kutta Method for Initial Value Problems with Rapidly Varying Right-Hand Sides. *ACM Transactions on Mathematical Software* **16**(1990), 201-222.
[3] Czech Statistical Office. https://www.czso.cz/. (last accessed 05.05.17.).
[4] Dostál, P.: *Soft computing v podnikatelství a veřejné správě.* Brno: CERM Akademické nakladatelství, 2015.
[5] Fehlberg, E.: Klassische Runge-Kutta-Formeln vierter und niedrigerer Ordnung mit Schrittweiten-Kontrolle und ihre Anwendung auf Waermeleitungsprobleme. *Computing* **6 (**1970), 61-71.
[6] Fiala, P.: *Dynamické vytváření cen* a *alokace zdrojů* v *síti*. Professional. Publishing, Praha, 2016.
[7] Griffiths W., E., Hill, R., C., and Judge G., G.: *Learning and Practicing Econometrics*. John Wiley & Sons. Inc., USA, 1992.
[8] Grytsay, I. V., Ghusainov, D. Y., and Půža, B.: Stabilization of System with Pure Delay in Linear Part. *Probleme of nonlinear analysis in engineering system*, Kazaň, 2007, **13**/(2007), 27-34.
[9] Gujarati, D., N.: *Basic Econometrics*. Fourth ed. (international ed.). McGraw-Hill Companies, USA, 2003.
[10] Holman, R.: *Ekonomie* . Nakladatelství C. H. Beck, Praha, 1999.
[11] Hušek, R.: *Ekonometrická analýza*. Praha: VŠE v Praze, Nakladatelství Oeconomica, 2007.
[12] Kiguradze, I., and Půža, B.: *Boundary value problems for systems of linear functional differential equations*, Folia Facul. Sci. Natur. Univ. Masar. Brun., Mathematica 12, Masaryk University, Brno, 2003.
[13] Kolmanovskii, V., B., Myshkis, A., D., and Nosov, V., R.: Modern Theory of Equations with Delay from the Point of View of its Applications. *Sovremennyje. Problemy MatFiz*., Tbilisi, 1987, 280–290.
[14] Krasovskij, N., N.: Optimal Processes in Systems with Time Lag. In: *Proceedings of the* 2nd IFAC–Congress, Basel, 1963, London, 1963, 327- 332.
[15] Maple Help System. http://www.maplesoft.com/support/help/Maple/view.aspx?path=dsolve/numeric/delay (last accessed 05.05.17.).
[16] Maplesoft. http://www.maplesoft.com/ (last accessed 05.05.17.).
[17] Marušiak, P., and Olach, R.: *Funkcionálne diferenciálne rovnice.* EDIS, Žilina, 2000.
[18] Pražák, P. Diferenciální rovnice a jejich použití v ekonomii. Dissertation work. Karlova univerzita, Matematicko- fyziální fakulta, Praha, 2006.
[19] Samuelson, P. A., and Nordhaus, W. D.: *Ekonomie*. Nakladatelství Svoboda, Praha, 1991.
[20] Shampine, L. F., and Corless, R. M.: Initial Value Problems for ODEs in Problem Solving Environments. *J. Comp. Appl. Math* **125**(2000), 31-40
[21] Simon, C. P., and Blume, L.: *Mathematics for Economists*. W. W. Norton & Company, London, New York, 1994.
[22] Smith, H.: *An Introduction to Delay Differential Equations with Applications to the Life Sciences*. Springer, New York Dordrecht Heidelberg London, 2011.
[23] Sterman, J, D.: *Buseiness Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill Companies, USA, 2000.
[24] Varian, H. R.: *Intermediate Microeconomics*. *A modern approach*. 8th ed. W. W. Norton & Company, Inc., London, New York, 2010.

Comparison of efficiency results for financial institutions using different DEA models

Lucie Chytilová[1], Jana Hančlová[2]

Abstract. The aim of the article is to analyze efficiency of financial institutions (mainly banks) from the Visegrad Group countries. The basic intermediation approach is used for the identification of input and output variables. Different types of conventional methods of data envelopment analysis (DEA) are used and compared. Advantages and disadvantages of each approach are discussed and analyzed for the area of the research.

Keywords: financial institutions, Data Envelopment Analysis, efficiency, Visegrad Group.

JEL classification: C61, G21, D24
AMS classification: 90C05

1 Intoduction

The situation for all kinds of financial institutions is very difficult all around the world. This paper mainly focuses on banking industry in the Central Europe and especially countries of the Visegrad Group. These countries had to keep up with a lot of changes - structural changes of banking regulation and financial market, development of security market, bigger stock market or new regulations by the European Union, see [4]. This all causes the big interest in the topic - efficiency in the banking sector of the Visegrad Group.

The Visegrad Group contains four countries - the Czech Republic, Hungary, Poland and Slovakia. These countries are not connected just geographically. The historical background is very similar in these countries. All four countries had been part of the Eastern Bloc which had fallen apart in 1989. This led them to integrate and cooperate. They started to create the market economy and have opened themselves to the rest of the world. The transformation of the banking system was an essential part of the transformation. In 2004 all countries of the Visegrad Group have jointed the European Union. Slovakia has joined the third stage of the European Monetary Union in 2009. There are even some differences. It could be assumed that their financial and banking systems should not show major differences. Identification of the potential differences is one of the objectives of this paper.

This paper mainly focuses on the identification of efficiency of commercial banks in the Visegrad Group during time period from 2005 to 2015 by using special models of DEA. More precisely, different models of DEA for missing data are used. The difference is in the idea how the missing data are treated. The reason of using these models is the general problem with datasets in this type of research. In this case, the dataset contained 35 banks from the Visegrad Group countries, but some variables were missing. To still be able to make the analysis, first the estimation of these missing variables was done. Then the special DEA model was used. Model was calculated for the time period and close analysis of all results were done for each bank as well as for each country. All calculations were done by GAMS software.

The rest of the paper has the following structure: Section 2 provides the formulations, DEA models and interval DEA models to estimate efficiency bounds for units with missing values. In Section 3, input and output variables are defined and information about the missing data and their estimation are given. Section 4 focuses on application and efficiency analysis of banks in Visegrad Group based on the dataset with missing values. The discussions about the results are provided in this section as well. Section 5 gives some conclusions and remarks.

2 Methodology

2.1 Classic DEA

Data Envelopment Analysis (DEA) is a non-parametric approach. It is widely used for measuring relative efficiency of decision making units (DMUs) with multiple inputs and outputs. Assume, there is a set of T DMUs (DMU$_k$ for $k = 1, ..., T$), let input and output variables data be $X = \{x_{ik}, i = 1, ...R; k = 1, ...T\}$ and $Y = \{y_{jk}, j = 1, ...S; k = 1, ...T\}$, respectively. Also, u_i for $i = 1, ...R$ and v_j for $j = 1, ..., S$ be the weights of the i^{th} input variable and the j^{th} output variable, respectively. Mathematically, the relative efficiency score of DMU$_k$ can be

[1]VŠB-TU Ostrava, Faculty of Economics, Sokolská 33, 702 00, Ostrava, Czech Republic, lucie.chytilova@vsb.cz
[2]VŠB-TU Ostrava, Faculty of Economics, Sokolská 33, 702 00, Ostrava, Czech Republic, jana.hanclova@vsb.cz

defined as:

$$e_k = \frac{\sum_{j=1}^{S} v_j y_{jk}}{\sum_{i=1}^{R} u_i x_{ik}}, \text{ for } k = 1, ..., T. \tag{1}$$

Charnes et al. [3] have proposed the following CCR model to measure the efficiency score of the under evaluation unit, DMU_Q where $Q \in \{1, ..., T\}$:

$$\begin{aligned} & \max e_Q = \frac{\sum_{j=1}^{S} v_j y_{jQ}}{\sum_{i=1}^{R} u_i x_{iQ}}, \\ \text{s.t.} \quad & \sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} \leq 0, \quad k = 1, ..., T, \\ & u_i \geq 0, \qquad\qquad\qquad\qquad i = 1, ..., R, \\ & v_j \geq 0, \qquad\qquad\qquad\qquad j = 1, ..., S. \end{aligned} \tag{2}$$

The model (2) is non-linear. It is the model of linear-fractional programming. The model (2) could be transferred by Charmes-Cooper transformation to the standard linear programming problem:

$$\begin{aligned} & \max e_Q = \sum_{j=1}^{S} v_j y_{jQ}, \\ \text{s.t.} \quad & \sum_{i=1}^{R} u_i x_{iQ} = 1, \\ & \sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} \leq 0, \quad k = 1, ..., T, \\ & u_i \geq 0, \qquad\qquad\qquad\qquad i = 1, ..., R, \\ & v_j \geq 0, \qquad\qquad\qquad\qquad j = 1, ..., S, \end{aligned} \tag{3}$$

where $Q \in \{1, ..., T\}$. DMU_Q is CCR-efficient if and only if $e^* = 1$ and if there exists at least one optimal solution $(\mathbf{u}^*, \mathbf{v}^*)$ with $\mathbf{u}^* > \mathbf{0}$ and $\mathbf{v}^* > \mathbf{0}$ for the set $Q \in \{1, ..., T\}$. The inefficient units have a degree of relative efficiency that belongs to interval $[0, 1)$. Note: The model must be solved for each DMU separately.

The model (3) is called a multiplier form of the input-orient-CCR model. However, for computing and data interpretation, it is preferable to work with model that is dual associated to model (3). The model is referred as envelopment form of input-oriented CCR model, see [3]. There also exists a multiplier form and envelopment form of output-oriented CCR model. Both models give the same results, see [3].

Banker et al. [2] have extended CCR model. The extended model is called BCC model and considers variable returns to scale assumption. The model has convex envelope of data which leads to more efficient DMUs. The mathematical model of dual multiplier form of input-oriented BCC model is:

$$\begin{aligned} & \max e_Q = \sum_{j=1}^{S} v_j y_{jQ} - v_0, \\ \text{s.t.} \quad & \sum_{i=1}^{R} u_i x_{iQ} = 1, \\ & \sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} - v_0 \leq 0, \quad k = 1, ..., T, \\ & u_i \geq 0, \qquad\qquad\qquad\qquad i = 1, ..., R, \\ & v_j \geq 0, \qquad\qquad\qquad\qquad j = 1, ..., S, \\ & v_0 \in (-\infty, \infty), \end{aligned} \tag{4}$$

where v_0 is the dual variable assigned to the convexity condition $\mathbf{e^T}\lambda = \mathbf{1}$ of envelopment form of input-oriented BCC model. Note: The BCC model can be rewritten into the envelopment form or changed into the output orientation.

The input-oriented BCC model will continue into the next section.

2.2 DEA for missing data

There exist many improvements of classical DEA models. In this paper, the main problems are the missing data. To deal with this issue the special model has to be calculated. This model is define below. As a inspiration, the model of Smirlis et al. [9] has been used. Their model has been done for the output-oriented DEA model, so the transformation was needed.

Assume, there are T DMUs, each using R input variables to produce S output variables. For any unit k ($k = 1, ..., T$), the level of its j^{th} output variable ($j = 1, ..., S$) is denoted by y_{jk} and by x_{ik} the level of its i^{th} input variable ($i = 1, ..., R$). Unlike the original DEA model, the interval DEA assumes that some of the crisp input x_{ik} and output y_{jk} variables are not known. For these variables is only known that they lie within bounded intervals, i.e. $x_{ik} \in [x_{ik}^L, x_{ik}^U]$ and $y_{jk} \in [y_{jk}^L, y_{jk}^U]$, with the upper and lower bounds of intervals x_{ik}^L, x_{ik}^U, y_{jk}^L, y_{jk}^U to be strictly positive constants.

To be able to introduce intervals, instead of the exact data, into the model (4), some transformation is needed.

Variables x_{ik} and y_{jk} are expressed in terms of new variables s_{ik} and p_{jk}, respectively, to convert the non-linear model to a linear one. These new variables locate the level of input and output variables within the bounded intervals $[x_{ik}^L, x_{ik}^U]$ and $[y_{jk}^L, y_{jk}^U]$, respectively, as follow:

$$\begin{array}{l} x_{ik} = x_{ik}^L + s_{ik}(x_{ik}^U - x_{ik}^L), \quad i = 1, ..., R;\ k = 1, ..., T \text{ with } 0 \le s_{ik} \le 1, \\ y_{jk} = y_{jk}^L + t_{jk}(y_{jk}^U - y_{jk}^L) \quad j = 1, ..., S;\ k = 1, ..., T \text{ with } 0 \le t_{jk} \le 1. \end{array}$$

Applying the above transformation to model (4), the following linear model is obtained:

$$\begin{array}{lll} & \max e_Q = \sum_{j=1}^{S} v_j \big(y_{jjQ}^L + t_{jkQ}(y_{jkQ}^U - y_{jkQ}^L)\big) - v_0, & \\ \text{s.t.} & \sum_{i=1}^{R} u_i \big(x_{ikQ}^L + s_{ikQ}(x_{ikQ}^U - x_{ikQ}^L)\big) = 1, & \\ & \sum_{j=1}^{S} v_j \big(y_{jj}^L + t_{jk}(y_{jk}^U - y_{jk}^L)\big) - \sum_{i=1}^{R} u_i \big(x_{ik}^L + s_{ik}(x_{ik}^U - x_{ik}^L)\big) - v_0 \le 0, & k = 1, ..., T, \\ & u_i \ge 0, 0 \le s_{ik} \le 1 & i = 1, ..., R, \\ & v_j \ge 0, 0 \le t_{jk} \le 1 & j = 1, ..., S, \\ & v_0 \in (-\infty, \infty). & \end{array} \tag{5}$$

It can be see that for input and output variables, there are new terms $u_i s_{ik}$ and $v_j t_{jk}$, respectively. These new terms may be replaced by new variables $q_{ik} = u_i s_{ik}$ and $p_{jk} = v_j t_{jk}$ which meet the needed of conditions. The model (5) can be rewritten as follows:

$$\begin{array}{lll} & \max e_Q = \sum_{j=1}^{S} \left(v_j y_{jjQ}^L + p_{jkQ}(y_{jkQ}^U - y_{jkQ}^L)\right) - v_0, & \\ \text{s.t.} & \sum_{i=1}^{R} \left(u_i x_{ikQ}^L + q_{ikQ}(x_{ikQ}^U - x_{ikQ}^L)\right) = 1, & \\ & \sum_{j=1}^{S} v_j y_{jj}^L + \sum_{j=1}^{S} p_{jk}(y_{jk}^U - y_{jk}^L) - \sum_{i=1}^{R} u_i x_{ik}^L - \sum_{i=1}^{R} q_{ik}(x_{ik}^U - x_{ik}^L) - v_0 \le 0, & k = 1, ..., T, \\ & q_{ik} - u_i \le 0 & i = 1, ..., R, \\ & p_{jk} - v_j \le 0 & j = 1, ..., S, \\ & u_i, v_j \ge 0 & \forall i, j \\ & q_{ik}, p_{jk} \ge 0 & \forall i, j, k \\ & v_0 \in (-\infty, \infty). & \end{array} \tag{6}$$

In model (6), unknown variables under estimation are weights u_i, v_j and new variables q_{ik}, p_{jk} that denote the level of input and output variables within the bounded intervals. For more details see [9].

3 Input and Output Variables

The important task for efficiency measurements is to identify the right and relevant variables for the calculation. There are known two main approaches for evaluation of financial institutions (as production units). The main difference between these two approaches is treatment of deposits. The production approach views deposits as output. Banks are producers of deposits, loans and other services. Inputs are define as physical variables. This approach was found by Benston [1]. Benston also found disadvantages - detailed database is required and it does not take into consideration the interest costs. The second approach was found by Sealy and Lindley [8] - intermediation approach. Banks are financial intermediaries between depositors and creditors. They collect deposits and other liabilities to apply them as interest-earning assets. Deposits are considered as input. In this case operating cost and interest cost are considered. It is the most common approach nowadays. This approach is used in this article as well.

The required data set of input and output variables have been collected from Bankscope[3]. In Table 3, it is seen more precise description of input and output variables. All variables, besides the number of employees, are in thousand Euros.

This paper deals with time period form 2005 to 2015. During this time period there have been many banks in the region. Not all of the banks have been at the market for whole period. Some banks have been closed or opened during the mentioned period. The assumption for DEA method is that decision making units (banks) have to be homogenous during whole analyzed time period. According to this assumption, selection of banks for the time period had to be done. There had been found 24 banks which had all needed input and output variables. On the other hand, there had been found 11 banks which were missing one or three variables in the time period

[3]https://bankscope.bvdinfo.com/

Variables	Description in the balance sheet	Unit of measurement
Input Variables		
Labour (x_1 - EMP)	Number of employees	Number
Physical capital (x_2 - FA)	Fixed assets = Tangible + Intangible assets	Thousands of Euro
Loanable funds (x_3 - DEP)	Deposits + Short term funding	Thousands of Euro
Output Variables		
Advances (y_1 - ADV)	Loans and advances to banks	Thousands of Euro
Investments (y_2 - SEC)	Other securities	Thousands of Euro
Non-interest income (y_3 - NEA)	Non-earning assets	Thousands of Euro

Source: own processing, 2017

Table 1 Description of input and output variables

(banks with more then three missing variables had been removed). As these missing variables can be replaced[4]. Finally, there had been detected 35 banks for the analysis. A list of specific 35 Visegrad banks is available from the author.Finally, there had been selcted 11 banks from the Czech Republic (CZ) and Poland (PL), 7 banks from Hungary (HU) and 6 banks from Slovakia (SK) for the whole time period.

4 Results and Discussions

In Table 2, there are seen average results for all banks during the whole time period. More precisely, there are results for four models and three types of categories. Four models in Table 2 are following:

- classic BCC model (4), where missing values are treating as series mean values - BCC_M;
- classic BCC model (4), where missing values are treating as trend values - BCC_T;
- classic BCC model (4), where missing values are values from regression analysis - BCC_R;
- special DEA model for interval data (6) - BCC_I.

The categories which are used for better analysis are following:

- description is overall average efficiency for each year of the period;
- size of banks are based on assets;
- origin of the bank.

Results for all banks can not be described according to the length of the paper.

Table 2 shows, according to overall average efficiency, models BCC_T and BCC_R have same results. Model BCC_M gives same (2007, 2008, 2009, 2010, 2012, 2013, 2014) or little bit different results. Different efficiency scores are generally smaller. Just in year 2012, it may be seen that efficiency score by model BCC_M is higher then efficiency score given by BCC_T and BCC_R models. This shows that different methods for replacing the missing values are causing some differences. The last model, let's say the special model, for interval data (BCC_I) gives results which are closer to BCC_M. BCC_I model is in range from 0.7352 to 0.8083. The lowest value of the technical efficiency is in year 2008. This is not surprising, as this is the year of the financial crises. This values is smaller then the average (0.7753) but the difference is small (0.04). It means that the technical efficiency is rather around 75% than around 80%.

Similar differences in results are not just seen for the over all average efficiencies. Average efficiencies by size and origin give similar results as well.

5 Conclusion

In this paper, the basic and special input-oriented BCC DEA models are calculated for banking industry in the Visegrad Group to obtain the efficiency score. The banking efficiency score is calculated for the time period from 2005 to 2015. First, the missing values have been estimated by several basic methods and then DEA basic models and DEA special model have been used. There have been found some differences between all used methods. Overall, the results have shown that the most efficient banking industry is in the Czech republic. The worst is in Poland (according to number of efficient banks which have been efficient for whole period) or in Slovakia (according to the efficiency score overall time). This gives similar results as the previous research (see 1. Chapter of [7]), where just 27 banks were used. Also, it has been seen that different size of bank is important factor as well.

For further research, the set of input and output variables should be extended or revised as well as the number

[4]Method for replacing missing values are: regression, trend and series mean value.

BCC_M											
	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
efficiency	0.7937	0.8156	0.7621	0.7380	0.8171	0.7449	0.7749	0.8125	0.7993	0.7425	0.7859
micro	0.9139	0.8568	0.7508	0.7934	0.8382	0.7197	0.8124	0.7933	0.8035	0.7817	0.8253
small	0.6821	0.8176	0.7882	0.6103	0.7870	0.6540	0.6370	0.7057	0.7134	0.6400	0.6983
medium	0.7344	0.6923	0.6100	0.6487	0.7240	0.6778	0.7390	0.7854	0.7376	0.6198	0.7260
big	0.8595	0.8959	0.9116	0.9307	0.9207	0.9550	0.9378	0.9872	0.9569	0.9609	0.9037
CZ	0.9394	0.8933	0.8573	0.8289	0.9011	0.7911	0.8403	0.8466	0.8389	0.8139	0.9738
HU	0.6800	0.6651	0.5713	0.5820	0.7658	0.7044	0.7996	0.8008	0.7777	0.7004	0.9071
PL	0.7508	0.8079	0.7690	0.7516	0.7592	0.8126	0.8048	0.9157	0.8995	0.8751	0.8021
SK	0.7729	0.8909	0.8454	0.7608	0.8427	0.6070	0.6007	0.6155	0.6083	0.4971	0.4323
BCC_T											
	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
efficiency	0.8012	0.8168	0.7621	0.7380	0.8171	0.7449	0.7737	0.8125	0.7993	0.7425	0.7863
micro	0.9187	0.8568	0.7508	0.7934	0.8382	0.7197	0.8124	0.7933	0.8035	0.7817	0.8253
small	0.6878	0.8225	0.7882	0.6103	0.7870	0.6540	0.6363	0.7057	0.7134	0.6400	0.6996
medium	0.7389	0.6923	0.6100	0.6487	0.7240	0.6778	0.7370	0.7854	0.7376	0.6198	0.7261
big	0.8748	0.8959	0.9116	0.9307	0.9207	0.9550	0.9355	0.9872	0.9569	0.9609	0.9037
CZ	0.9466	0.8976	0.8573	0.8289	0.9011	0.7911	0.8403	0.8466	0.8389	0.8139	0.9738
HU	0.6950	0.6651	0.5713	0.5820	0.7658	0.7044	0.7989	0.8007	0.7777	0.7004	0.9072
PL	0.7569	0.8079	0.7690	0.7516	0.7592	0.8126	0.8032	0.9157	0.8995	0.8751	0.8021
SK	0.7733	0.8909	0.8454	0.7608	0.8427	0.6070	0.5981	0.6155	0.6083	0.4971	0.4334
BCC_R											
	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
efficiency	0.8012	0.8168	0.7621	0.7380	0.8171	0.7449	0.7737	0.8125	0.7993	0.7425	0.7863
micro	0.9187	0.8568	0.7508	0.7934	0.8382	0.7197	0.8124	0.7933	0.8035	0.7817	0.8253
small	0.6878	0.8225	0.7882	0.6103	0.7870	0.6540	0.6363	0.7057	0.7134	0.6400	0.6996
medium	0.7389	0.6923	0.6100	0.6487	0.7240	0.6778	0.7370	0.7854	0.7376	0.6198	0.7261
big	0.8748	0.8959	0.9116	0.9307	0.9207	0.9550	0.9355	0.9872	0.9569	0.9609	0.9037
CZ	0.9466	0.8976	0.8573	0.8289	0.9011	0.7911	0.8403	0.8466	0.8389	0.8139	0.9738
HU	0.6950	0.6651	0.5713	0.5820	0.7658	0.7044	0.7989	0.8007	0.7777	0.7004	0.9072
PL	0.7569	0.8079	0.7690	0.7516	0.7592	0.8126	0.8032	0.9157	0.8995	0.8751	0.8021
SK	0.7733	0.8909	0.8454	0.7608	0.8427	0.6070	0.5981	0.6155	0.6083	0.4971	0.4334
BCC_I											
	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
efficiency	0.7935	0.8075	0.7595	0.7352	0.8083	0.7375	0.7691	0.8065	0.7899	0.7383	0.7825
micro	0.9138	0.8548	0.7416	0.7890	0.8380	0.7163	0.8124	0.7933	0.7989	0.7782	0.8116
small	0.6821	0.8173	0.7877	0.6103	0.7771	0.6466	0.6356	0.7057	0.7091	0.6400	0.6994
medium	0.7344	0.6649	0.6100	0.6422	0.7061	0.6762	0.7358	0.7661	0.7095	0.6075	0.7255
big	0.8587	0.8959	0.9113	0.9307	0.9142	0.9355	0.9162	0.9808	0.9569	0.9609	0.9028
CZ	0.9394	0.8913	0.8570	0.8252	0.8950	0.7882	0.8403	0.8466	0.8312	0.8086	0.9738
HU	0.6800	0.6351	0.5713	0.5820	0.7507	0.7020	0.7964	0.7973	0.7642	0.6929	0.9072
PL	0.7502	0.8079	0.7612	0.7516	0.7478	0.7941	0.7908	0.9071	0.8858	0.8749	0.7904
SK	0.7729	0.8909	0.8450	0.7506	0.8427	0.6038	0.5966	0.6030	0.6060	0.4930	0.4330

Table 2 Average efficiencies for all types of models

of DMUs. This extension can provide better analysis (connection to size or origin). Moreover, a different version of DEA model may be established in the future.

Acknowledgements

This article was supported by Student Grant Competition project (SP2017/114) of Faculty of Economics, VŠB-Technical University of Ostrava and the Czech Science Foundation through project No. 17-23495S.

References

[1] Benston, G. J.: Branch Banking and Economies of Scale. *Journal of Finance* **20** (1965), 312–333.

[2] Banker, R. D., Charnes, A., and Cooper, W. W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* **30** (1984), 1078–1092.

[3] Charnes, A., Cooper, W. W., and Rhodes, E. Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.

[4] Fang, Y., Hasan, I., Marton, K. and, Waisman, M.: Bank valuation in new EU member countries. *Economic Systems* **38** (2014), 55–72.

[5] Gulati, R., and Kumar, S.: Impact of non-traditional activities on the efficiency of Indian banks: an empirical investigation. *Macroeconomics and Finance in Emerging Market Economies* **4** (2011), 125–166.

[6] Hančlová, J., and Chytilová, L.: The impact of inclusion of non-traditional activities on efficiencies in European banking industry using the CCR-I model. In: *Proceedings of the Conference on Strategic Management and its Support by Information Systems 2015*, VŠB-TUO, Ostrava, 2015, 208–220.

[7] Hančlová, J., et al. *Optimization Problems in Economics and Finance.* VŠB-TUO, Ostrava, 2015.

[8] Lyroudi, K., and Angelidis, D.: Measuring Banking Productivity of the Most Recent European Union Member Countries. *Journal of Economics and Business* **9** (2006), 37–57.

[9] Smirlis, Y., G., Maragos, E., K., and Despotis D., K. Data envelopment analysis with missing values: An interval DEA approach. *Applied Mathematics and Computation* **177** (2006), 1–10.

Benchmarking of countries at Summer Olympic Games using two-stage DEA models

Josef Jablonský[1]

Abstract. After important sports events as the Summer Olympic Games are, the participating countries are ranked according to the number of gold, silver and bronze medals. A lexicographic ranking is usually applied which leads to higher ranking of countries with one gold and no other medals comparing to countries without any gold but with several silver or bronze medals. Moreover, this ranking does not take into account the specific conditions of the countries (population, economic strength that is usually measured by GDP per capita, etc.). The aim of the paper is not only to evaluate the absolute achievements of the countries but evaluate their performance with respect to the resources they can use. A two-stage DEA model is formulated and solved by an original SBM procedure. The first stage evaluates the performance of the countries in training of athletes and the second stage evaluates the achievements of the nominated athletes. The models with variable returns to scale and weight restrictions are applied. The models and their results are illustrated on the case of Olympic Games 2016 and compared with results given by traditional models.

Keywords: data envelopment analysis, two-stage model, ranking, Olympic Games.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Ranking of countries after important sport events belongs to the most significant results with respect to performance and success of nations at the event. Summer or Winter Olympic Games are probably the most influencing and biggest meetings of athletes. They are organized in four-year cycles and triples of medals in many contests (gold, silver, bronze) are competed. Traditional ranking officially published by the International Olympic Committee is based on lexicographic method, i.e. the countries with a higher number of golds are ranked higher regardless on the number of medals on lower levels. This leads to hardly acceptable results – e.g. a country with only one gold and no other medals is better than a country with no gold but several silvers and/or bronzes. Moreover, the ranking based only on absolute numbers of medals is questionable because it do not take into account resources of the countries (population, economic strength, tradition in sports, etc.).

Regardless on taking or non-taking into account additional variables the problem of ranking is a typical multiple criteria decision making (MCDM) problem that can be solved using one of the MCDM methods. The aim of this paper is to verify non-traditional approaches for this problem. Data envelopment analysis (DEA) models evaluate relative performance of decision making units (DMUs) according to the defined input and output variables and there have been published many studies dealing with the same or similar subject. This paper applies two-stage DEA models for assessing the performance of countries and compares the results with traditional DEA models and lexicographic ranking.

Similar studies have already been published. Novelty of this paper consists in selection of variables and application of an original DEA two-stage model. Lozano et al. [12] analyzed results of last five Summer Olympic Games using DEA model with variable returns to scale and two inputs (population and GDP) and three outputs (gold, silver and bronze medals). Weight restrictions have been used in order to assign higher importance to golds than to silvers and the same for silvers to bronzes. An alternative ranking is suggested in Lins et al. [11]. This ranking is based on ability of each country to win medals in relation to its available resources. A DEA model based on the premise that the sum of the gains is zero (constant sum of outputs) is developed and its results are discussed. Churilov and Flitman [4] have demonstrated that traditional measures frequently reported lack the necessary descriptive power. The productivity measurement DEA approach is used for producing a new ranking of participating teams at Olympic Games in Sydney 2000. The study of Li et al. [9] extends previous DEA studies by incorporating multiple sets of nation-specific assurance regions into the DEA model and derives ranking of countries for six past Olympic Games. Wu et al. [15] and [16] have analyzed using cross-efficiency models for ranking of countries and

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, Praha 3, e-mail: jablon@vse.cz.

extend this kind of models by new features. Zhang et al. [17] discuss the issue of preferences in DEA models for measuring the performance of nations at the Olympic Games and propose a new DEA model with lexicographic preference to measure the performance of the nations. Shirouyehzad and Yazdani [13] use an output-oriented BCC model with the number of male and female athletes, received medals in two previous Olympic as the inputs and gold, silver and bronze medals are the only outputs of the model. This model is solved for female and male athlete combination and their separation. According to our knowledge Li et al. [10] is the most recent paper dealing with ranking of nations at Olympic Games and first time uses two-stage DEA model for this purpose.

The rest of the paper is organized as follows. Section 2 contains basic formulation of the problem and definitions of DEA models further used in the study. Two-stage models and their modification for purposes of the study are formulated in Section 3. Final section informs about numerical experiments and discusses the results given by all models.

2 Formulation of the problem and theoretical background

The problem of ranking countries according to the number of gold, silver and bronze medals they won is a typical multiple criteria decision making problem that can be solved using available MCDM methods. SAW, TOPSIS, AHP/ANP, PROMETHEE class methods, and ELECTRE class methods belong to the most often applied ones. Their using is not possible (except the simplest ones) without appropriate software tools. Unfortunately, there are just few packages that allow solving MCDM problems using various methods simultaneously. One of them is Sanna that is designed as add-in MS Excel application and can be downloaded from author's web pages. Sanna covers 6 MCDM methods (except other functionalities) including the above mentioned ones. A brief information about this system can be found e.g. in Jablonsky [6]. Another MCDM application is IZAR – it is a stand-alone system with similar performance as Sanna – a more detailed information about IZAR is in Kalcevova and Fiala [7]. Real-world applications of MCDM methods in many areas including sports are numerous. Except this there are many attempts to analyze typical MCDM problems using DEA models – one of the interesting application is described in Smirlis et al. [14].

The problem of ranking of countries according to the number of medals is a MCDM problem with three criteria. If we want to solve this problem using a DEA model the three criteria can be taken as only outputs and a DEA model without explicit outputs can be used. Nevertheless, this analysis does not take into account specifics of the countries – it is clear that a country with higher population, more developed country or a country with higher past tradition in sports has a higher chance to reach better results. These characteristics can be taken as inputs and finally the efficiency score of each country can be derived using a DEA model. In our analyses below we will work with radial models under the assumption of variable returns to scale (VRS). A formulation of a VRS output-oriented DEA model is given below.

Let us suppose that the set of decision making units (DMUs) contains n elements. The DMUs are evaluated by m inputs and r outputs with input and output values x_{ij}, $i = 1,2,\ldots,n$, $j = 1,2,\ldots,m$ and y_{ik}, $i = 1,2,\ldots,n$, $k = 1,2,\ldots,r$, respectively. The efficiency of the q-th DMU can be expressed as the weighted sum of outputs divided by the weighted sum of outputs with weights reflecting the importance of single inputs/outputs v_i, $i = 1,2,\ldots,m$ and u_k, $k = 1,2,\ldots,r$. Banker et al. [2] have introduced the linearized for VRS output-oriented DEA model as follows:

$$\text{Minimize} \quad \theta_q = \sum_{j=1}^{m} v_j x_{qj} + \nu$$

subject to (1)

$$\sum_{k=1}^{r} u_k y_{qk} = 1,$$

$$\sum_{k=1}^{r} u_k y_{ik} - \sum_{j=1}^{m} v_j x_{ij} - \nu \le 0, \; i = 1,2,\ldots,n,$$

$$u_k, v_i \ge \varepsilon, \qquad k = 1,2,\ldots,r, i = 1,2,\ldots,m,$$

$$\nu \text{ - free},$$

Where $\theta_q \ge 1$ is the efficiency score of the unit under evaluation and ε is an infinitesimal constant.

It is clear that the importance of gold medals is higher that silver ones and they are more important than bronze medals. In order to ensure this relation the following sets of constraints is added to model (1):

$$u_k \ge \alpha_k u_{k+1}, \qquad k = 1,2,\ldots,r-1, \tag{2}$$

$$u_k - u_{k+1} \ge u_{k+1} - u_{k+2}, \qquad k = 1,2,\ldots,r-2,$$

where α_k, $k = 1, 2,\ldots, r-1$, is a constant that defines relation between pairs of weights. Model (1)-(2) can be used for calculation of efficiency scores of countries and they can be ranked according to their values. The units identified as efficient have identical maximum efficiency score $\theta_q = 1$. Then they can be ranked using models for ranking of efficient units – one of the options is Andersen and Petersen model [1] that removes the q-th constraint from the set of constraints of model (1). The efficiency scores of the originally efficient units is then lower than 1 which allows their ranking.

3 Two-stage serial DEA models

A general serial two-stage process is defined as follows. The inputs of the first stage are transformed into its outputs and all or at least some of these outputs are used as inputs of the second stage that are spent for production of final outputs. Let us denote the input values of the first stage x_{ij}, $i = 1,2,\ldots,n$, $j = 1,2,\ldots,m$ and the output values of the first stage that serve as inputs for the second stage as z_{il}, $i = 1,2,\ldots,n$, $l = 1,2,\ldots,p$. Final outputs of the second stage (final outputs of the production process) are y_{ik}, $i = 1,2,\ldots,n$, $k = 1,2,\ldots,r$. We do not suppose that there are any inputs entering the second stage independently on the first stage, and no outputs leaving the first stage without any connection to the second one.

The problem of evaluation of countries according to their results at Olympic Games can be formulated in a more general way as a two-stage process (Figure 1). The first stage contains two inputs (Population and GDP per capita) and one output (Team size). This stage measures efficiency of the country with respect to sports education and ability to train high-quality athletes. The second stage evaluates efficiency of the team that was nominated for the Games.

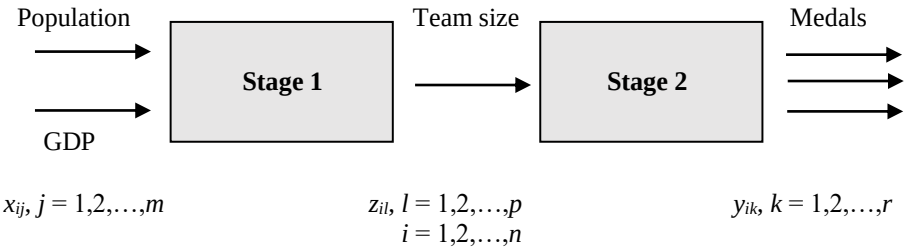


Figure 1 Two-stage serial model

There have been formulated several models for efficiency evaluation of two-stage processes. Kao and Hwang [8] and Chen et al. [3] belong to the most often known. Their formulation can be found either in original papers [3] and [8] or in Jablonsky [5]. Except these two models an original SBM model derived from Chen model is applied. Its mathematical formulation is the following:

Minimize

$$\psi_q = \frac{1-\frac{1}{m}\sum_{j=1}^{m}\left(s_j^- / x_{qj}\right)}{1+\frac{1}{r}\sum_{k=1}^{r}\left(s_k^+ / y_{qk}\right)} \tag{3}$$

subject to

$$\sum_{i=1}^{n} x_{ij}\lambda_i + s_j^- = x_{qj}, \qquad j = 1, 2, ..., m, \tag{4}$$

$$\sum_{i=1}^{n} z_{il}\lambda_i \geq \tilde{z}_{ql}, \qquad l = 1, 2, ..., p, \tag{5}$$

$$\sum_{i=1}^{n} z_{il}\mu_i \leq \tilde{z}_{ql}, \qquad l = 1, 2, ..., p, \tag{6}$$

$$\sum_{i=1}^{n} y_{ik}\mu_i - s_k^+ = y_{qk}, \qquad k = 1, 2, ..., r, \tag{7}$$

$$(1-\tau)z_{ql} \leq \tilde{z}_{ql} \leq (1+\tau)z_{ql}, \; l = 1, 2, ..., p, \tag{8}$$

$$\lambda_i \geq 0, \; \mu_i \geq 0, \; s_k^+ \geq 0, \; s_j^- \geq 0,$$

Where s_j^-, $j = 1, 2,\ldots, m$, are slack variables belonging to the inputs of the first stage, s_k^+, $k = 1, 2,\ldots, r$, are surplus variables belonging to the final outputs, τ is the parameter that fixes the rate of positive and negative deviations of values z_{ql} and $\tilde{z}_{ql}$ and ψ_q is the total efficiency score of the q-th unit. Constraints (4)-(7) correspond to the same set of constraints in Chen model. Constraints (8) ensure that the variable $\tilde{z}_{ql}$ differs from the original value of the l-th output of the first stage by τ100% at the most. The objective function of the model is the ratio of average relative slacks in the input space and average relative surplus variables in the output space. The unit under evaluation is efficient if all slack/surplus variables equal 0, i.e. the optimal objective function of the model is $\psi_q^* = 1$. Lower values indicate lower efficiency. Model (3)-(8) is not linear in its objective function but can be transformed into linear program and then, using its dual, constraints (2) can be applied in order to discriminate among importance of final outputs. Moreover, convexity constraints for λ and μ variables can be added in order ensure variable returns to scale assumption.

Country	Gold	Silver	Bronze	Population	GDP	Team size
United States	46	37	38	321.419	15094.00	554
Great Britain	27	23	17	65.138	2431.59	366
China	26	18	26	1371.22	7298.10	413
Russia	19	18	19	144.097	1857.77	265
Germany	17	10	15	81.413	3570.56	425
Japan	12	8	21	126.958	5867.15	338
France	10	18	14	66.808	2773.03	395
South Korea	9	3	9	50.617	1116.25	205
Italy	8	12	8	60.802	2194.75	309
Australia	8	11	10	23.781	1371.76	421
Netherlands	8	7	4	16.936	836.26	242
Hungary	8	3	4	9.845	140.03	160
Brazil	7	6	6	207.847	2476.65	465
Spain	7	4	6	46.418	1490.81	306
Kenya	6	6	1	46.05	33.62	89
Jamaica	6	3	2	2.726	15.07	68
Croatia	5	3	2	4.224	63.85	87
Cuba	5	2	4	11.389	60.81	120
New Zealand	4	9	5	4.595	130.68	199
Canada	4	3	15	35.852	1736.05	314
Uzbekistan	4	2	7	31.299	45.36	70
Kazakhstan	3	5	9	17.544	186.20	104
Colombia	3	2	3	48.229	331.65	147
Switzerland	3	2	2	8.287	635.65	104
Iran	3	1	4	79.109	86.53	64
Greece	3	1	2	10.824	298.73	95
Argentina	3	1	1	43.417	445.99	213
Denmark	2	6	7	5.676	332.68	122
Sweden	2	6	3	9.799	538.13	152
South Africa	2	6	2	54.957	408.24	137
Ukraine	2	5	4	45.198	165.25	203
Serbia	2	4	2	7.098	45.04	104
Poland	2	3	6	37.999	514.50	243
North Korea	2	3	2	25.155	22.00	35
Belgium	2	2	2	11.285	511.53	108
Thailand	2	2	2	67.959	345.65	54
Slovakia	2	2	1	5.424	95.99	51
Georgia	2	1	4	3.679	14.37	39
Azerbaijan	1	7	10	9.651	63.40	56
Belarus	1	4	4	9.513	55.14	121
Turkey	1	3	4	78.666	773.09	103
Czech Republic	1	2	7	10.551	215.22	105

Table 1 Data set – characteristics of the countries[2]

[2] www.medalspercapita.com [accessed March 15, 2017]

4 Numerical experiments

The data set for the analysis is presented in Table 1. This table contains 42 countries (selection of the best countries at Olympic Games 2016 in Rio) with a positive number of medals of all kinds. Except the number of medals the table contains population of the country in millions of people, GDP per capita in thousands of USD, and team size at the Games. The countries in Table 2 are ranked in lexicographic order according to the number of gold, silver and bronze medals respectively. We have applied various DEA models, all under the assumption of VRS and output-orientation. In models with medals as outputs constraints (2) with $\alpha = 2$ are added. Efficiency scores and ranking for selected models are given in Table 2:

- Model 1 – model without explicit inputs and three outputs (G, S, B);
- Model 2 – Stage 1 – two inputs (population, GDP), one output (team size);
- Model 3 – Stage 2 – one input (team size), three outputs (G, S, B);
- Model 4 – two-stage model (Figure 1); Kao and Hwang model;

	Country	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
1	United States	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)
2	Great Britain	1.675 (2)	1.189 (15)	1.103 (6)	1.316 (4)	1.500 (9)	1.325 (7)
3	China	1.769 (3)	1.219 (17)	1.322 (10)	1.598 (6)	2.678 (36)	1.340 (8)
4	Russia	2.260 (4)	1.662 (27)	1.071 (5)	1.796 (7)	1.523 (10)	1.142 (6)
5	Germany	2.706 (5)	1.046 (11)	2.081 (19)	2.215 (12)	2.317 (23)	2.428 (13)
6	Japan	3.482 (7)	1.378 (21)	2.115 (20)	2.928 (17)	3.318 (41)	2.245 (11)
7	France	3.289 (6)	1.107 (14)	2.338 (21)	2.596 (15)	2.245 (22)	2.362 (12)
8	South Korea	5.111 (10)	1.831 (31)	1.920 (17)	3.611 (23)	1.865 (16)	3.116 (20)
9	Italy	4.607 (8)	1.401 (23)	2.553 (24)	3.601 (21)	2.364 (26)	2.742 (16)
10	Australia	4.625 (9)	1.000 (1)	3.507 (31)	3.515 (20)	1.873 (17)	3.312 (22)
11	Netherlands	5.609 (11)	1.344 (19)	2.427 (23)	3.292 (19)	1.810 (15)	3.451 (24)
12	Hungary	5.751 (12)	1.254 (18)	1.697 (12)	2.213 (11)	1.499 (8)	3.349 (23)
13	Brazil	6.435 (13)	1.000 (1)	5.392 (40)	5.401 (29)	2.322 (24)	5.398 (39)
14	Spain	6.572 (14)	1.391 (22)	3.656 (33)	5.170 (28)	2.333 (25)	4.489 (32)
15	Kenya	7.167 (15)	1.040 (10)	1.107 (7)	1.197 (3)	1.166 (4)	3.184 (21)
16	Jamaica	7.667 (17)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)
17	Croatia	9.200 (20)	1.427 (24)	1.497 (11)	2.147 (9)	1.457 (6)	2.945 (18)
18	Cuba	9.200 (19)	1.057 (12)	2.056 (18)	2.205 (10)	1.473 (7)	3.519 (25)
19	New Zealand	7.588 (16)	1.000 (1)	2.679 (25)	2.726 (16)	1.637 (12)	2.941 (17)
20	Canada	8.000 (18)	1.355 (20)	4.509 (38)	6.143 (32)	2.744 (37)	5.318 (38)
21	Uzbekistan	10.963 (23)	1.544 (25)	1.283 (9)	2.138 (8)	1.432 (5)	1.817 (10)
22	Kazakhstan	9.548 (21)	2.009 (32)	1.736 (14)	3.601 (22)	2.196 (21)	1.807 (9)
23	Colombia	15.333 (29)	1.599 (26)	3.986 (37)	6.538 (34)	2.546 (32)	4.978 (36)
24	Switzerland	15.334 (30)	2.327 (36)	2.889 (27)	6.850 (36)	2.601 (33)	4.358 (30)
25	Iran	15.334 (30)	2.398 (37)	1.704 (13)	4.347 (26)	2.084 (20)	3.046 (19)
26	Greece	15.334 (32)	2.411 (38)	2.741 (26)	6.737 (35)	2.534 (31)	5.839 (40)
27	Argentina	15.334 (33)	1.199 (16)	5.978 (41)	7.512 (38)	2.672 (35)	13.970 (42)
28	Denmark	10.963 (23)	1.735 (29)	2.353 (22)	4.185 (25)	2.034 (19)	2.429 (14)
29	Sweden	12.870 (25)	1.706 (28)	3.441 (30)	6.022 (31)	2.441 (28)	4.221 (29)
30	South Africa	12.900 (26)	1.815 (30)	3.119 (29)	5.790 (30)	2.395 (27)	4.663 (34)
31	Ukraine	13.455 (27)	1.011 (8)	4.869 (39)	4.983 (27)	2.456 (29)	4.960 (35)
32	Serbia	16.125 (34)	1.027 (9)	2.915 (28)	3.108 (18)	1.731 (13)	3.912 (27)
33	Poland	14.800 (28)	1.102 (13)	6.432 (42)	7.150 (37)	2.664 (34)	6.454 (41)
34	North Korea	18.429 (35)	2.205 (34)	1.000 (1)	2.567 (14)	1.755 (14)	1.000 (1)
35	Belgium	21.143 (38)	2.473 (39)	3.961 (36)	10.195 (41)	3.179 (40)	4.979 (37)
36	Thailand	21.143 (38)	4.398 (42)	1.863 (16)	9.062 (40)	2.993 (39)	2.665 (15)
37	Slovakia	21.500 (42)	3.165 (41)	1.813 (15)	6.264 (33)	2.480 (30)	3.614 (26)
38	Georgia	21.143 (38)	1.000 (1)	1.242 (8)	1.488 (5)	1.115 (3)	1.000 (1)
39	Azerbaijan	10.571 (22)	2.313 (35)	1.000 (1)	2.472 (13)	1.553 (11)	1.112 (5)
40	Belarus	18.500 (36)	1.000 (1)	3.936 (35)	4.041 (24)	1.984 (18)	4.507 (33)
41	Turkey	21.143 (38)	3.048 (40)	3.807 (34)	11.980 (42)	3.452 (42)	4.189 (28)
42	Czech Rep.	19.733 (37)	2.039 (33)	3.625 (32)	7.627 (39)	2.746 (38)	4.459 (31)

Table 2 Results of the models – efficiency scores and ranking

- Model 5 – two-stage model (Figure 1); Chen model (geometric average of efficiency scores of the first and second stages);
- Model 6 – output oriented two-stage SBM model (3)–(8) - $\tau = 0.05$.

The results of the models presented in the previous section can be discussed and analyzed in a more detail. Ranking given by Model 1 corresponds more or less to original official ranking (first column of Table 2). This is clear because this model takes into account just the number of medals and no other characteristics. Model 2 measures appropriateness of the team size with respect to the country resources. Models 3 evaluates performance of the Olympic teams with respect to the number of medals they won. There are very big differences in ranking according these two models – this is apparent at most for countries as Australia, Canada, Belarus, North Korea, Azerbaijan and others. The last three models are two-stage models and they have the same assumptions – VRS, weight restrictions and the same set of variables. All these models lead to the result that fully efficient countries are USA and Jamaica but there are significant differences especially in case of Kao and Hwang model comparing to the other two ones.

The advantage of the SBM two-stage model consists in its independence on orientation of the model. Its properties and relation to other models must be further investigated which opens space for future research in this field.

Acknowledgements

The research is supported by the Czech Science Foundation, project no. 16-01821S.

References

[1] Andersen, P., and Petersen, N.C.: A procedure for ranking efficient units in data envelopment analysis. *Management Science*, **39** (1993), 1261–1264.
[2] Banker, R.D., Charnes, A., and Cooper, W.W., Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, **30** (1984), 1078–1092.
[3] Chen, Y., Liang, L., and Zhu, J.: Equivalence in two-stage DEA approaches. *European Journal of Operational Research*, **193** (2009), 600–604.
[4] Churilov, L., and Flitman, A.: Towards fair ranking of Olympics achievements: the case of Sydney 2000. *Computers & Operations Research*, **33** (2006), 2057–2082.
[5] Jablonsky, J.: Data envelopment models with network structure. In: *Proceedings of the International Conference Mathematical Methods in Economics*, 2012, 367–372.
[6] Jablonsky, J.: MS Excel based software support tools for decision problems with multiple criteria. *Procedia Economics and Finance*, **12** (2014), 251–258.
[7] Kalcevova, J., and Fiala, P.: IZAR – Multicriteria Decision Support System. In: *Proceedings of the International Conference Mathematical Methods in Economics*, 2006, 277–282.
[8] Kao, C., and Hwang, S.N.: Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *European Journal of Operational Research*, **185** (2008), 418-429.
[9] Li, Y., Liang, L., Chen, Y., and Morita, H.: Models for measuring and benchmarking Olympics achievements. *Omega*, **36** (2008), 933–940.
[10] Li, Y., Lei, X., Dai, Q., and Liang, L.: Performance evaluation of participating nation at the 2012 London Summer Olympics by a two-stage data envelopment analysis. *European Journal of Operational Research*, **243** (2015), 964–973.
[11] Lins, M.P.E., Gomes, E.G., Soares de Mello, J.C., and Soares de Mello, A.J.: Olympic ranking based on a zero sum gains DEA model. *European Journal of Operational Research*, **148** (2003), 312–322.
[12] Lozano, S., Villa, G., Guerrero, F., and Cortés, P.: Measuring the performance of nations at the Summer Olympics using data envelopment analysis. *Journal of the Operational Research Society*, **53** (2002), 501–511.
[13] Shirouyehzad, H., and Yazdani, F.: Performance evaluation and ranking of participation Asian countries in 2012 London Olympic Games through data envelopment analysis. *Journal of Data Envelopment Analysis and Decision Science*, **3** (2014), 1–11.
[14] Smirlis, Y., Despotis, D., Jablonsky, J., and Fiala, P.: Identifying „Best-Buys" in the Market of Prepaid Mobile Telephony: an Application of Imprecise DEA. *International Journal of Information Technology & Decision Making*. **3** (2004), 167–177.
[15] Wu, J., Liang, L., and Yang, F.: Achievement and benchmarking of countries at the Summer Olympics using cross efficiency evaluation method. *European Journal of Operational Research*, **197** (2009), 722–740.
[16] Wu, J., Liang, L., and Chen, Y.: DEA game cross-efficiency approach to Olympic rankings. *Omega*, **37** (2009), 909–918.
[17] Zhang, D., Li, X., Meng, W., Liu, W.: Measuring the performance of nations at the Olympic Games using DEA models with different preferences. *Journal of the Operational Research Society*, **60** (2009), 983–990.

A Comparative Analysis of the Information Society in Poland and Selected Countries

Anna Janiga-Ćmiel[1]

Abstract. The aim of the study was to test contagion in selected countries of the European Union in the years 2000-2016. The transmission channel to be examined was a variable denoting Internet usage by people aged 16 -74. Taxonomic methods were adopted to select countries for the purpose of comparison. The paper presents a model of Internet usage in a particular country under study, which was built using first order linear differential equations.

Keywords: differential equation, taxonomic, information society, contagion effect.

JEL Classification: C01, C51.
AMS Classification: 91G70, 91B84

1 The information society – an overview of the problem

The emergence of the Internet and the ongoing development of technologies exert a powerful impact on the world, national economies, society, culture, as well as human relations and everyday life. The achievements of modern civilization are a sign of not only positive but also negative changes taking place in society as we are becoming heavily dependent on them. Nowadays, information resources are in constant and widespread use, which leads to our daily dependence on technological solutions. The widening of the access to the Internet and modern technologies is referred to as informatization (computerization) and a direct effect of this process is the emergence of the information society. The term was coined in 1963 by Tado Umesamo to describe the evolution of the society based on information industries. In Europe, it was first used in 1978, and in the USA – in the 1980s [2].

The cycle of the information society evolution is often compared to Toffler's concept of three waves, each representing a stage in the transition process: the agricultural society, the industrial society and the information society. The subject literature offers a wide variety of definitions of the information society. One of them defines the information society as "a social system, developed as a result of the modernization processes, in which information systems and information resources determine the employment structure and an increase in social wealth (national income), and provide a basis for the civilization's orientation" [2].

We are dominated by the computer and Internet, which on the one hand is a useful communication tool, but on the other, may lead to alienation.

An easy access to the Internet allows for instant verification of information and provides learning opportunities. In addition, it enables greater flexibility of various kinds of activities, which can be done not only from home, but from any place in the world, e.g. managing a company, shopping, taking care of personal finance by means of mobile banking, watching films on telephones, etc [8, 17]. However, it also poses numerous threats, the most common being susceptibility to information distortion, reduced privacy due to increased control, job losses resulting from the implementation of automation and robotics solutions, the emergence of new types of crime. Nevertheless, it has to be noted that the Internet provides free access to databases around the world. At present, more than 70% of computer networks making up the Internet are located in the USA and thus that is the place where the global database network is the most strongly concentrated. Even in the countries which are deeply attached to their traditions and culture, such as Japan, modern technologies have contributed to the disappearance of the old world. A lot of countries make every effort to fight against digital exclusion, which is embedded in social exclusion. In conclusion, the Internet has brought about irreversible changes, both positive and negative, to the contemporary economic and social life.

2 A model of trends in Internet usage over time

During the course of the analysis, a selected group of people from a particular EU country was investigated to find out whether they have or do not have access to a computer. The following variables were introduced into the model: *BK*- lack of permanent access to a computer, *MK*- permanent access to a computer. Two groups were distinguished among the people who lack permanent access to a computer: people who have no Internet access

[1] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, 1 Maja 50, 40-287 Katowice, Poland, anna.janiga-cmiel@ue.katowice.pl.

at all (*BMKI*) and those with occasional access (*MKI*). Also, the people who have permanent access to a computer were divided into two groups: one included those who have an opportunity to use the Internet, while the other was made up of people who do not have Internet access. For each group symbols *MI* and *BI* were used respectively. Finally, within the *BI* group, people who do not want to have any Internet access were identified. They were described by means of variable *OI*. *N* was used to denote the population size.

The values of variables presented above are expressed as a percentage. The variable derivative shows its growth $N \cdot MI'(t+1)$, i.e. an annual increase in the number of people with permanent access to the Internet, $N \cdot BI'(t+1)$- an annual increase in the number of people with no Internet access. It is predicted that $N \cdot MI'(t+1)$ will be positive values, and $N \cdot BI'(t+1)$ – negative values. Also, an appropriate equation is used to investigate an increase in the number of people who declare that they do not want to have any Internet access – this constitutes the third equation of the model determining $N \cdot OI'(t+1)$ dynamics. In all equations, the model parameters are denoted by symbols *a,b,c,d,e,f*. The model parameters are calculated relative to the size of relevant subpopulations. Thus, an increase in the group of people with permanent Internet access depends on both the percentage of people having permanent access to a computer and permanent access to the Internet in the preceding years, and the percentage of people who lack computer access but seek Internet access and manage to gain it occasionally:

$$N \cdot MI'(t+1) = a \cdot MK(t) \cdot MI(t) + b \cdot BK(t) \cdot MKI(t) \tag{1}$$

Another group taken into account consists of people who do not have Internet access – this percentage value depends on the number of people who lack permanent access to a computer as well as those who do not have occasional access to a computer. The dependency is represented by equation:

$$N \cdot BI'(t+1) = c \cdot BK(t) \cdot BMKI(t) + d \cdot BI(t) \tag{2}$$

The people who have no Internet access, are not interested in it and do not intend to acquire any devices with Internet access are represented in the model by equation:

$$N \cdot OI'(t+1) \tag{3}$$

The size of this group is affected by the percentage of $BI(t)$, i.e. people who lack permanent access to the Internet, and the $BMKI(t)$ group, i.e. people who do not have even sporadic access to the Internet.

$$N \cdot OI'(t+1) = f \cdot BMKI(t) + e \cdot BI(t). \tag{4}$$

Thus, the initial pattern of dependencies occurring in the development of Internet usage in society is presented by the following dynamic model:

$$\begin{cases} N \cdot MI'(t+1) = a \cdot MK(t) \cdot MI(t) + b \cdot BK(t) \cdot MKI(t) \\ N \cdot BI'(t+1) = c \cdot BK(t) \cdot BMKI(t) + d \cdot BI(t) \\ N \cdot OI'(t+1) = f \cdot BMKI(t) + e \cdot BI(t) \end{cases} \tag{5}$$

The left-hand side of the equations is expressed as an increment in the variable for the next year $(t+1)$; it is described by the values of the explanatory variables from the previous year t and is approximated by replacing the increment in the variable by its derivative. For example, for equation (1) we obtain:

$$N \cdot [MI(t+1) - MI(t)] \approx N \cdot MI'(t+1). \tag{6}$$

The same procedure applies to the other equations.

3 A study of correlation coefficients

The crises of the 1990s and of the 21st century were mainly financial ones, connected with problems in the banking sector. They indicted the weakness of the financial systems and involved loss of liquidity and currency devaluation. The underlying causes of these crises were investigated by numerous scientists, including M. K. Brunnermeier, P. O. Gourinchas, S. S. Skrzypek [16]. The research led to the realization that during the crisis there was an increase in the correlation between stock markets, as well as bond markets and currency markets around the world. So, "a considerable growth in a correlation among various financial markets due to shocks is called a contagion effect" [5, 6]. Theoretical and empirical studies strive to explain all kinds of causes, mechanisms and sources of contagion. The World Bank proposes three definitions of contagion: broad, restrictive and a very restrictive one (World Bank), for example: Contagion is defined as the cross-country transmission of shocks or the general cross-country spillover effects [4, 10,11].

The concept of the contagion effect is associated with the transmission channel [4, 10,11, 14]. The channel is understood as "a cross-country link between two or more national economies, through which there can happen mutual transmission of economic impulses, particularly various economic crises, no matter what kind of link it is or how it manifests itself, as long as it produces cross-country effects". The World Bank indicates four kinds of links which can constitute transmission channels: real, financial, behavioral and political. It is extremely difficult to examine and analyse the causes of contagion due to a shortage of information about the transmission channels. Another difficulty lies in the interconnections between the transmission channels. Crises often spread because investors in various countries hold the same assets.

The occurrence of a contagion effect can be analysed by means of various methods, both mathematical and econometric. Each of these methods is based on its own assumptions and offers different possibilities of measuring the contagion effect [6, 7]. There is ongoing discussion among researchers which approach and method best describes the problem. Two alternative approaches can be found in the literature. Some scientists, e.g. B. Eichengreen, R. Glick and A.K. Rose, analyze the transmission of contagion. Others, including P. Masson, K. Forbes and R. Rigobon, I. Pritsker focus on the source of contagion [5, 7].

The most commonly adopted method is a correlation analysis, proposed by Forbes and Rigobon [6], but there are also other ones, such as the VAR models, used by C. Favero and F. Giavazzi. The subject literature comprises also an event study analysis, presented by S. Lizondo, C. Reinhart [6, 10, 11] or a hidden factor analysis, developed by A. Ng and R. Hodrick, G. Corsetti, M. Dungey, M. Pericoli, M. Sbracia. A number of other methods can be found in Masson's works and the ones of G. Marais, C. Cailleteau, D. Diamond, P. Dybving [14], M. Fratzscher [5, 6, 10, 11].

In the article the occurrence of the contagion effect is verified by means of the Forbes and Rigobon test. In the case of dramatic variance changes, one can identify periods of low and high variance, which is referred to as heteroskedasticity. The formula for bias correction of the correlation coefficient devised by the researchers takes into account variance changes. In the method the following denotations are introduced [5, 10, 11]: σ_L^2 - the low variance, σ_H^2 - the high variance.

For the period of low variation we have correlation coefficient ζ_L, whereas correlation coefficient ζ_H is related to high variation.

In the next step of the analysis, an adjusted correlation coefficient is introduced, where the correction factor is a relative gain in the variation, represented by formula (7) proposed by Forbes and Rigobon. [5,6]

$$\beta = \frac{\sigma_H^2 - \sigma_L^2}{\sigma_L^2} \tag{7}$$

ζ_H denotes a correlation coefficient calculated for the period when the variance was assumed as high σ_H^2. The adjusted correlation coefficient ζ_S is given by (8) and covers the whole period, where ζ_H is a coefficient for the entire period before adjusting.

$$\zeta_S = \frac{\zeta_H}{\sqrt{1+\beta-\beta\cdot\zeta_H^2}} \tag{8}$$

Where respectively:

ζ_L - a correlation coefficient for the period of low variance,

ζ_H - a coefficient of correlation for the period of high variance, ζ_S - an adjusted correlation coefficient.

The correction factor β in the formula (8) assumes non-negative values, wherein the value of zero indicates that there were no shocks in the development of the phenomena under study. However, the value significantly higher than zero is a sign of the occurrence of shocks. [6, 10, 11]. Now, it is necessary to verify the relevance of the difference in coefficient ζ_S and coefficient ζ_L before the contagion period, i.e. during the period of low variance. When testing the existence of contagion, we form the following hypothesis:

$$H_0 : \zeta_S = \zeta_L \qquad H_1 : \zeta_S > \zeta_L \tag{9}$$

According to Forbes and Rigobon, the equality of adjusted correlation coefficient ζ_S with correlation coefficient ζ_L, prior to the moment of contagion, has to be verified. An alternative hypothesis predicts that the contagion effect occurs.

4 An example – a model of Internet usage

The paper presents an analysis of a particular set of diagnostic features. It spans the years 2000-2016 and is based on data for 24 selected countries of the European Union published by the Central Statistical Office as well as national and OECD statistical yearbooks.

The variables, whose relevance and statistical accuracy were first verified, formed a basis for the classification of the countries according to Internet usage by physical persons aged 16-74. In the first stage of the analysis a taxonomic method was employed [9]. The equal-intensity matching was chosen to investigate the level of Internet usage in Poland, select countries that would represent the groups of countries with a higher level of the examined phenomenon, and order them all according to the Internet usage intensity indicator. The equal-intensity matching belongs to direct methods [15]. It involves grouping of the analyzed objects by verifying a statistical hypothesis in order to find out whether the feature under study shows a statistically significant difference across the compared objects or not [3, 12, 15].

Three groups of homogeneously developed countries were identified and one country was selected to represent each group. The first, the most developed group in terms of Internet usage, was represented by Luxembourg, the second group – by Poland, and the third one, the least developed – by Romania [12].

A basis for the construction of a model was thus established, i.e. a model was built for the country representing each group. The structural parameters of the model were estimated using the classical least square method (CLS). The model was pre-verified.

The following model is obtained for Poland:

$$\begin{cases} MI(t+1) = 0.2615e^{0.04(526t^2-324t+162)} + 0.0515e^{0.02(321t-216)} \\ BI(t+1) = -0.0462e^{-0.02(124t^2-92t+6)} - 0.00006e^{-0.06(14t+5)} \\ OI(t+1) = -0.00515e^{0.01(5t^2+6)} - 0.00005e^{0.0712(12t+6)} \end{cases} \tag{10}$$

In order to estimate the model, the adopted variables were expressed as the number of users per 1,000 people. In the last stage of the analysis the model was solved using methods for solving first order linear non-homogeneous differential equations, while taking into account the initial conditions [13].

The model for Luxembourg takes the form:

$$\begin{cases} MI(t+1) = 0.8641e^{0.07(926t^2+256t+8.1)} + 0.4326e^{0.04(526t+11)} \\ BI(t+1) = -0.1012e^{-0.04(214t^2+184t+125)} - 0.00017e^{-0.08(25t+38)} \\ OI(t+1) = -0.0012e^{0.11(24t^2+51t+34)} - 0.00011e^{0.171(t+32)} \end{cases} \tag{11}$$

The model for Romania is as follows:

$$\begin{cases} MI(t+1) = 0.0114e^{0.02(212t^2-101t+14)} + 0.0014e^{0.002(11t-5)} \\ BI(t+1) = -0.0103e^{-0.01(104t^2-151t+1.5)} - 0.00001e^{-0.01(5t+3)} \\ OI(t+1) = -0.0051e^{0.004(3t+1.03)} - 0.0002e^{0.003(2t+1.03)} \end{cases} \tag{12}$$

In each analyzed case, the solution was obtained in the exponential form as a solution of the linear differential equation and a sum of the general and particular solutions.

5 A study of contagion

The last stage of the study attempted to analyze the occurrence of contagion when the Internet is used by people aged 16-17 in selected EU countries (Poland, Denmark, Luxembourg, France, United Kingdom, Czech Republic, Slovakia, Romania, Ukraine). The values of correlation coefficients of the examined variable for Poland and the selected countries were determined taking into account a delay of *t* periods backwards.

In addition, based on the previous broader study, the analysis covered one non-EU country – Ukraine. For Poland we obtain:

Country	Poland/Denmark	Poland/Luxembourg	Poland/France	Poland/United Kingdom	Poland/Czech Republic	Poland/Slovakia	Poland/Ukraine	Poland/Romania
ζ_S	0.25	0.12	0.24	0.2	0.16	0.14	0.11	0.15
ζ_L	0.79	0.92	0.63	0.7	0.55	0.49	0.08	0.28

Table 1 The values of coefficients ζ_S ζ_L

The comparison of all the values according to procedure (Forbes and Rigobon test) shows that in all the cases except for Ukraine the inequality $|\zeta_S| < |\zeta_L|$ is fulfilled. The values of all coefficients obtained during the course of the analysis are positive, thus confirming an upward trend in the number of users in particular countries. In none of the countries except Ukraine was evidence of contagion found.

4 Conclusion

To sum up, in all the countries under study a steady annual increase in the number of Internet users can be observed. The percentage of people who do not want to use the Internet has been decreasing. The findings of the study are confirmed by the signs of the model parameter estimates, positive indicating a rise while negative a fall.

During the contagion analysis positive correlation coefficients were obtained, which also indicates a growth in the number of Internet users in particular countries. In the case of Poland and the selected EU countries, no contagion was observed with regard to the examined variable. The only situation when weak contagion was detected was the cross-country analysis of Poland and Ukraine.

References

[1] Alvarez, F., and Barlevy, G.: *Mandatory Disclosure and Financial Contagion*. NBER Working Paper, 21328, MA 02138, Cambridge, 2015.
[2] Bliźniuk G., and Nowak J., S.: *Społeczeństwo informacyjne 2005*. PTI, Katowice, 2005.
[3] Chomątowski S., and Sokołowski A.: *Taksonomia struktur*. Przegląd Statystyczny. R. XXV - Zeszyt 2, 1978.
[4] Dach, Z.: *Polityka makroekonomiczna w warunkach kryzysu i jej wpływ na gospodarkę.* Wolters Kluwer, Warszawa, 2011.
[5] Fiszeder, P., and Razik W.,: *Analiza efektu zarażania na przykładzie zależności pomiędzy indeksem WIG a indeksami wybranych rynków akcji na świecie*. Acta Universitatis Nicolai Copernici, Ekonomia XXXIII, Toruń, 2003.
[6] Fiszeder, P.: *Modele klasy GARCH w empirycznych badaniach finansowych*. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń, 2009.

[7] Forbes, K., and Rigobon R.: No Contagion, Only Interdependence*: Measuring Stock Market Co-movments*, Journal of Finance, 57, 2002.
[8] Goban-Klas T., and Sienkiewicz P. *Społeczeństwo informacyjne. Szanse, zagrożenia, wyzwania*. Wyd. Fundacji Postępu Telekomunikacji, Kraków, 1999.
[9] Hadaś-Dyduch, M., *Predicting change in the labor market in global aspect,* The 10-th International Days of Statistics and Economics, Conference Proceedings, 8-10.09. 2016, Prague, Czech Republic, Libuse Macakova, Melandrium, 2016, 543-554.
[10] Janiga-Ćmiel, A.: *Detecting Shocks in The Economic Development Dynamics of Selected Countries.* Folia Oeconomica Stetinensia, 13 (2), 2013, 120-133.
[11] Janiga-Ćmiel, A.*: An Analysis of Conditional Dependencies of Covariance Matrices for Economic Processes in Selected EU Countries.* Folia Oeconomica Stetinensia, 16 (2), 2016, 119-134.
[12] Janiga-Ćmiel, A.: *Analiza społeczeństwa informacyjnego wybranych krajów*. Studia Ekonomiczne nr 301, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, 2016, 94-103.
[13] Janiga-Ćmiel, A.: The application of stochastic equations to predict purely random phenomena. In: *Business entities in the face of contemporary economics*, (Sokół, A., Drab-Kurowska, A., and Budziewicz-Guźlecka, A., eds). Kartprint, Bratislava, 2017, 44-54.
[14] Naoui, K., Khemiri, S. and Liouane, N.: *Crises and Financial Contagion: The Subprime Crisis*. *Journal of Business Studies Quaterly*, *2* (1), 2010.
[15] Pociecha J., Podolec B., Sokołowski A., and Zając K.: *Metody taksonomiczne w badaniach społeczno-ekonomicznych*. PWN, Warszawa, 1988.
[16] Skrzypek, S.S.: Banki centralne wobec światowego kryzysu. In: *Banki centralne wobec kryzysu ekonomicznego.,* (Osiński, J., eds). Oficyna Wydawnicza SGH, Warszawa, 2010.
[17] Zorska A.: *Chaos czy twórcza destrukcja? Ku nowym modelom w gospodarce i polityce*. Oficyna Wydawnicza SGH, Warszawa, 2011.

Avoiding overfitting of models: an application to research data on the Internet videos

Radim Jiroušek [2,1], Iva Krejčová[2]

Abstract. A *model overfitting* is a well-known phenomenon both in statistics and machine learning. In this paper, we study this problem from the perspective of information theory. In this context, data-based model learning can be viewed as a transformation process; a process transforming the information contained in data into the information represented by a model. The overfitting of a model often occurs when one considers an unnecessarily complex model, which usually means that the considered model contains more information than the original data. Thus, using one of the basic laws of information theory saying that any transformation cannot increase the amount of information, we get the basic restriction laid on models constructed from data: A model is acceptable if it does not contain more information than the input data file.
This idea is also in agreement with the *minimum description length* principle that, roughly speaking, advices to prefer models described with a small number of parameters to more complex models.

Keywords: Data-based learning, probabilistic models, composition, information theory, MDL principle, overfitting, lossless encoding.

JEL classification: C52
AMS classification: 90B60

1 Introduction

It is perhaps unnecessary to explain in detail what is understood by overfitted models in AI [1] and/or in statistics [14]. Just recall that the notion is connected with models constructed from data, in particular, with the models reflecting noninformative properties of the source data files (like noise and other random properties that each randomly generated data file possesses). This phenomenon is often illustrated on two stochastically dependent variables, the dependence of which is linear. Because the dependence is stochastic, if randomly generated data are plotted in a graph, the respective dots are concentrated along a straight line describing the dependence. Naturally, only a part of them lies on the line. If one tries to find a curve that connects all the dots in the plot (see Fig. 1), the model is much less informative and cannot be used for prediction (neither for interpolation nor for extrapolation). It is important to realize that such a complex curve must be described (defined) by a much larger number of parameters than the straight line, which can be determined just by two points.

Going back to the ideas of von Mises [12] and Kolmogorov [9], who both explored relations interconnecting randomness, complexity, and information, we can learn that they were interested (among others) about "the quantity of information conveyed by an individual object 'x' about another individual object 'y' " [9]. Having two sequences $\mathcal{S}_1$, $\mathcal{S}_2$ of 0's and 1's, which are both lossless encoding of a considered model $\mathbf{M}$, we can thus deduce that both these sequences $\mathcal{S}_1, \mathcal{S}_2$ convey the same amount of information about model $\mathbf{M}$. The same holds true also for an optimum lossless encoding $\mathcal{S}^*$ of model $\mathbf{M}$. Since the mutual information between two objects is always less or equal the information contained in any of these objects, the length of encoding $\mathcal{S}^*$ is the best lower estimate of the amount of information contained in model $\mathbf{M}$ we have. Assuming also that there is no (relative) redundancy in sequence $\mathcal{S}^*$, we can take its length as an estimate of the amount of information (measured in bits) contained in model $\mathbf{M}$ (in what follows we will omit the word estimate, and will speak about the information, or amount of information contained in $\mathbf{M}$).

The above-presented ideas are independent of the type of considered models. The best model containing all the information contained in data is the respective data file itself. Therefore, using the above ideas, the amount of information in data equals the number of bits necessary to store an optimum lossless encoding of the respective data file. This enables us to compare the amount of information contained in data and that contained in a datalearned model. In case we get a model with a greater amount of information than that in data, we are sure, that some undesirable information has been added into the model. In addition to this, we know that regardless the way the data were collected, they always contain some specific part of the information, employment of which results in the overfitting of the model. It should not be included in the model. Therefore, all the considered models should contain less information than the input data. Thus we enforce a principle under which models with the amount

[1]Inst. of Information Theory and Automation, Acad. of Sciences, Prague, Czech Republic, radim@utia.cas.cz
[2]Faculty of Management, University of Economics, Jindřichův Hradec, Czech Republic, iva.krejcova@gmail.com

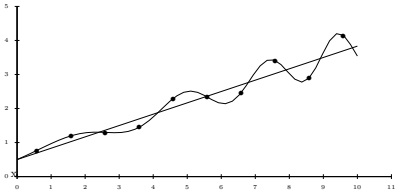


Figure 1 Overfitted linear dependence

of information greater or equal than that of the input data file are *unacceptable*. In fact, we accept only models containing *substantially* less information than the input data file. The meaning of the word *substantially* is usually left to the user's discretion.

Notice, that the above-mentioned principle is also fully compatible with the famous *Minimum Description Length* (MDL) principle that is often used in the process of model learning. For example, it was proposed for Bayesian network learning by Lam and Bacchus [11], (for general sources of this principle see e.g. [3], and [4]). For more discussion on this relationship see also [8].

The goal of this paper is to illustrate the above-described principle with the data-based learning of compositional models. Thus, in the next section, we introduce a minimum quota of concepts from compositional model theory necessary to describe these models and their lossless encodings. Section 3 is devoted to an example in which the approach is applied to data from a commercial research project *Video on Internet* performed by Nielsen Admosphere.

2 Compositional Models and their Encodings

By a compositional model, we understand a multidimensional probability distribution composed of a system of low-dimensional distributions (usually, marginals of the considered multidimensional one). From the viewpoint of this paper it is important to say that we consider discrete (finite-valued) variables $N = \{X_1, X_2, \ldots, X_n\}$; $\mathbb{X}_i$ denote the nonempty finite set of values of variable X_i. For a probability distribution $\pi(N)$, its marginal distributions for variables $M \subseteq N$ will be denoted either $\pi(M)$, or $\pi^{\downarrow M}$, and $\mathbb{X}_M$ denote the state-space of all the combinations of values of the included variables.

The most important concept from the theory of compositional models is an operator of composition $\triangleright$ that from two distributions, say $\kappa(K)$ and $\lambda(L)$ ($K, L \subseteq N$), constructs a more-dimensional distribution (denote it μ) of variables $K \cup L$

$$\mu(K \cup L) = \kappa(K) \triangleright \lambda(L) = \frac{\kappa(K) \cdot \lambda(L)}{\lambda(K \cap L)}. \tag{1}$$

In this paper, we do not need to care about the fact that the composition is not always defined, let alone what are its theoretical properties. For this, the reader is referred to survey papers [6, 7]. Here, we only have to know what are the parameters uniquely defining a compositional model.

Definition 1. Distribution $\pi(N)$ is a *compositional model* if there exists a cover $K_1, K_2, \ldots, K_m$ of N (i.e., $K_1 \cup \ldots \cup K_m = N$), such that[1]

$$\pi(N) = \pi^{\downarrow K_1} \triangleright \pi^{\downarrow K_2} \triangleright \ldots \triangleright \pi^{\downarrow K_m}. \tag{2}$$

It means that to define a compositional model it is enough to determine an ordered sequence of (marginal) distributions. Looking for its optimum lossless encoding we have to encode for each marginal distribution $\pi^{\downarrow K_i}$ the set of variables K_i and the respective probabilities. Realize that, while probability distribution $\pi(N)$ is defined with the help of $\prod_{i=1}^{n} |\mathbb{X}_i| - 1$ probabilities, to properly define compositional model (2) we need to specify only its marginals $\pi^{\downarrow K_i}$, i.e. we have to specify $\sum_{k=1}^{m} \left(\prod_{i \in K_k} |\mathbb{X}_i| - 1\right)$ probabilities.

It is perhaps unnecessary to say that speaking about an optimum encoding of models is an unattainable idealization. In practical situations, solving such an optimization problem would be intractable and therefore we have

[1]The operator of composition is not associative, the expression (2) is evaluated from left to right, i.e.,

$$\pi^{\downarrow K_1} \triangleright \pi^{\downarrow K_2} \triangleright \ldots \triangleright \pi^{\downarrow K_m} = \Big(\ldots \big((\pi^{\downarrow K_1} \triangleright \pi^{\downarrow K_2}) \triangleright \pi^{\downarrow K_3} \big) \triangleright \ldots \triangleright \pi^{\downarrow K_{m-1}} \Big) \triangleright \pi^{\downarrow K_m}.$$

to accept a suboptimal solution. This problem was studied in [8], where five different encoding approaches were compared. It was shown that none of the proposed approaches dominated others. So, for practical applications, we suggest the following solution: Consider a battery of encoding procedures and define the amount of information contained in data/model equal to the minimal length from all the binary encodings achieved by the considered approaches. For the sake of simplicity, we use in this paper only two encoding approaches: Direct Data Encoding which appears to be felicitous for the considered input data file encoding, and Huffman Lexicographic Encoding apt for the encoding of models (more precisely for the encoding of its building stones - marginal distributions). Before describing these encoding approaches, let us stress that it is of great importance to properly select the precision with which the respective probabilities are specified. This is because, as we will see below, the encoding of probabilities takes the substantial part of the whole code (in particular for Huffman Lexicographic Encoding).

Taking into account the precision, with which the probabilities are specified, is fully sensible also from the statistical point of view. Due to the accepted principle, the less data we have, the less amount of bits we may use to encode the model. It means, among others, that for small data files we cannot consider probability values specified with a high precision. This fully corresponds with the fact that having a small number of data, the confidence intervals for the estimates of probability parameters are rather wide. Therefore it does not have the sense to specify these estimates with a high precision, with a great number of digits.

To make the description of the encoding procedures as simple as possible, assume that all the values of all considered variables are nonnegative integers: $\mathbb{X}_i = \{0, 1, \ldots, |\mathbb{X}_i| - 1\}$.

Direct Data Encoding. We will use this type of encoding to encode a source data file. Consider a record $(x_1, x_2, \ldots, x_n)$ from a data file. It means that $x_i \in \mathbb{X}_i = \{0, 1, \ldots, |\mathbb{X}_i| - 1\}$. Therefore, each record can be unambiguously represented by the integer[2]

$$\left(\sum_{k=1}^{n-1} x_k \prod_{j=k+1}^{n} |\mathbb{X}_i| \right) + x_n,$$

which is an nonnegative integer less than $|\mathbb{X}_N| = \prod_{j=1}^{n} |\mathbb{X}_i|$, and therefore it can be encoded into[3] $\log_2 \lceil |\mathbb{X}_N| \rceil$ bits.

Huffman Lexicographic Encoding. To encode a low-dimensional probability distribution we have to specify the ordered sequence of its arguments (i.e., the respective variables), and the respective probabilities in a predefined ordering. To control the precision, with which the probabilities are specified, we determine probabilities of a low-dimensional distribution as a ratio $\frac{\ell}{base}$, where $base$ is a properly determined positive integer. The highest reasonable precision is to set $base$ equal to the size of a source data file. Decreasing this integer we decrease the precision of the specified probabilities. Setting $base = 1000$ means that we take all the probability estimates with three digits of precision. Rounding these estimates to two decimal digit means to consider $base = 100$. Nevertheless, it is important to realize that we can consider any number less or equal size of the data file at our disposal. Specifying the integer $base$ for the respective low-dimensional distribution, the probabilities are determined by integers, numerators of the respective ratios. To encode these integers we will apply the famous Huffman's encoding algorithm [5] (the reader not familiar with this encoding technique will see its application in Section 3).

Before proceeding to the illustrative example of the application of the proposed principle, let us remark that for the considered compositional models there are principally two ways of model simplification, which can be applied when getting unacceptably complex models:

- *Structure simplification* means to consider smaller sets of variables K_i.
- *Considering lower precision* of probabilities means to decrease the constant base (which should always be considered especially for distributions with higher dimensions).

[2]Such an integer is in fact an order number of the combination $(x_1, x_2, \ldots, x_n)$ in the ordering

$$(0, \ldots, 0, 0), (0, \ldots, 0, 1), \ldots, (0, \ldots, 0, \mathbb{X}_n - 1), (0, \ldots, 1, 0), \ldots, (\mathbb{X}_1 - 1, \ldots, \mathbb{X}_{n-1} - 1, \mathbb{X}_n - 1).$$

[3]$\lceil r \rceil$ denotes the smallest integer, which is not less than r.

3 Example

Because of the lack of space, we cannot describe the model learning principles in more details. Perhaps the best way how to clarify not-well-declared steps is to describe the process by an example. Therefore, in this section we consider a compositional model learning from data acquired in the framework of a commercial research *Video on the Internet* realized in 2016 by Nielsen Admosphere. This research project was oriented to answer questions concerning what and how often the respondents watch different types of videos on the Internet. The data were collected from 1207 respondents from among the Czech Internet population in the age of 15-83. From a large number of questions, for the purpose of this paper, we selected only 24 variables concerning demographic characteristics of respondents (5 variables: *age, sex, education, region,* and *size-of-municipality*), what type of electronic device they use (variables: $Q1$ - $Q7$), frequency and type of watched videos (variables $V1$ - $V6$, $T1$ - $T6$). Among them[4]

9 variables are dichotomic (2-valued) variables: $sex, Q1, \ldots, Q7, T6$,
9 variables are trichotomic (3-valued) variables: $edu, region, size, V1, \ldots, V6$,
6 variables have 4 values: $age, T1, \ldots, T5$.

Having these 24 variables, each combination of their values, i.e., each record from the input data file, can be uniquely encoded as a nonnegative integer less than

$$2 \times 4 \times 3 \times 3 \times 3 \times \underbrace{2 \times \ldots \times 2}_{7} \times \underbrace{3 \times \ldots \times 3}_{6} \times \underbrace{4 \times \ldots \times 4}_{5} \times 2 = 41\,278\,242\,816.$$

Among the 1207 records of the input data file only 8 records occur twice; it means that there is no way to find a substantially more efficient encoding of the data file than that encoding each record as a 36-digit binary number[5]. Therefore, we start the model learning process with an initial limitation given by the fact that the data file "contains the information" of $1207 \times 36 = 43\,452$ bits.

In the data-based process of model learning, we look for groups of variables defining the model (more precisely, the respective model is defined by an ordering of distributions defined for these groups of variables). Naturally, the goal is to group together variables that are highly interconnected. The strength of the interconnection is measured by the *measure of interdependence* (sometimes also called *information content*), which simplifies to the well-known mutual information for two-dimensional distributions [13]

$$MI(\pi(N)) = \sum_{(x_1,\ldots,x_n)\in\mathbb{X}_N} \pi(X_1 = x_1, \ldots, X_n = x_n) \log_2 \left(\frac{\pi(X_1 = x_1, \ldots, X_n = x_n)}{\pi(X_1 = x_1) \cdot \ldots \cdot \pi(X_n = x_n)} \right).$$

There are several strategies how to select an appropriate system of variable groups. It is beyond the scope of this paper to discuss them. What is important that we get groups forming a cover of the considered set of variables (each variable is included at least in one group), and that the measure of interdependence for variables in each group is high. The length of an efficient encoding of all the respective marginal distributions defines the "amount of information" contained in the model, which should be kept, in agreement with the introduced principle, sufficiently below the limit given by the "amount of information" contained in the data file.

Let us explain on the example of five-dimensional distribution how we compute the length of its binary encoding using the Huffman Lexicographic Encoding. Consider five-dimensional distribution of the demographic variables $\pi(age, sex, edu, region, size)$. The size of the respective five-way frequency table is $4 \times 2 \times 3 \times 3 \times 3 = 216$. When going to encode such a marginal distribution, first what we have to do is to choose the precision with which the probabilities will be specified, i.e., to specify the above mentioned constant $base$. Let us consider the highest reasonable precision for the given data file by setting $base = 1207$. Then we get the five-way frequency table (it is too big to present it here), in which we can see that the most frequent entry is 0; it occurs 36-times. The highest entry is 42, which, not surprisingly, appears in the table only once. The whole summary of values (frequencies) appearing in the considered five-way frequency table is in Table 1. Using a block code (fixed length code) for numbers from 0 to 42 we would need six bits to encode each entry. However, since some entries are much more frequent than others, famous Huffman variable length code [5] is more advantageous. It is known that this code is in a way optimal, and it assigns shorter code-words to more frequent entries. Moreover, the reader can see from Table 1 that the resulting code is a *prefix-free* code, which means that no code-word is a prefix of another code-word. To compute the length of the five-way frequency table encoding we need the lengths of individual code words. Thus to encode the frequency table corresponding to $\pi(age, sex, edu, region, size)$ we need

$$3\times(36+28+24+23)+4\times(17+15+13+8)+5\times(8+6+5+5+5)+6\times(4+3)+7\times(2+2)+8\times(\underbrace{1 \times \ldots \times 1}_{12}) = 856$$

[4]For the sake of simplicity, we clustered the values of variables to decrease the total number of values. For example, T and V variables have originally 6 and 7 values, respectively.

[5]$36 = \lceil \log_2(41\,278\,242\,816) \rceil = \lceil 35, 3 \rceil$

frequency	number of instances	Huffman code	length of code	frequency	number of instances	Huffman code	length of code
0	36	100	3	15	2	1011110	7
2	28	110	3	19	2	1011111	7
1	24	000	3	16	1	11110100	8
3	23	010	3	17	1	11110101	8
5	17	1010	4	18	1	11110110	8
4	15	1110	4	20	1	11110111	8
7	13	0010	4	24	1	11111000	8
6	8	0110	4	25	1	11111001	8
8	8	10110	5	27	1	11111010	8
11	6	00111	5	30	1	11111011	8
10	5	00110	5	32	1	11111100	8
12	5	01110	5	33	1	11111101	8
13	5	01111	5	39	1	11111110	8
9	4	101110	6	42	1	11111111	8
14	3	111100	6				

Table 1 Frequencies and their Huffman code

bits. Therefore, using Huffman code we need (in average) just $856/216 = 3.96$ bits to encode one entry of the considered five-way frequency table, i.e., one probability of the respective marginal probability distribution. Notice that we do not need to encode the respective denominator *base* because it can be got as a sum of all numbers from the frequency table. But we must not forget that we have to encode the respective Huffman code (otherwise nobody could decode it!). It means we have to encode three columns of Table 1: frequency, length of code, and Huffman code. For this we need:

11 bits to encode the maximum frequency;
29×6 bits to encode the frequencies;
6 bits to encode the maximum length of code word;
29×6 bits to encode the lengths of code words;
175 bits to encode the codewords.

Adding another 30 bits necessary to encode variables, for which the marginal distribution is defined, the described encoding needs 1426 bits to fully represent probability distribution $\pi(age, sex, edu, region, size)$. If we chose lesser precision of probabilities, say $base = 500$ (or $base = 200$), which means that we consider probabilities with the precision of 0.002 (0.005) we would need only 965 (550) bits to represent this distribution. Notice however that, if we considered five-dimensional marginal for four-valued variables $\pi(T1, \ldots, T5)$ with the highest reasonable precision $base = 1207$ we would need approximately four kilobits. Taking into account the fact that all reasonable models for the considered 24-dimensional distribution consisted of 18 – 21 marginal distributions, one can see that incorporating such space demanding marginals into the model would lead to models from our point of view unacceptable. Thus, for example, all five-dimensional distributions contained in the constructed *acceptable* models had always at least one binary variable among their arguments.

Recall that in the previous paragraph, we saw two ways how to keep learned models reasonably small: either we have to keep a simple structure, which means for compositional models that we keep the dimension of the considered marginals limited, or that we decrease the precision of probabilities. The trade-off between these two possibilities is fully under the control of the user. At this stage of research, we do not have any heuristics that could help them.

4 Conclusions

To avoid the well-known phenomenon of overfitting, different techniques for model verification and testing are usually used; starting with splitting the data into two parts: one part used for model learning and the other one for its testing (which decreases the amount of teaching data, though), or popular cross-validation technique. The common drawback of all these approaches is that they hardly ever make possible to distinguish the situations of overfitting from insufficient learning due to the lack of data (also called underfitting). Naturally, we do not claim the described principle is a general technique solving all such problems, but it can be used as one of stopping rules in the process of model learning. Its simplicity guarantees it can be used practically in all machine learning

tasks. And what is important, its careful application makes the machine learning specialists realize which parts of the models are space demanding. In connection with probabilistic models, it is also important to realize that simplification can be achieved (at least) in two different ways: by simplification of the structure of constructed models, or by the roughening of the estimates of probabilistic parameters.

Acknowledgements

The authors want to express their gratitude to Nielsen Admosphere, a.s. for providing the research data free of charge. The research was partially supported by GAČR under Grant No. 15-00215S.

References

[1] Berka, P.: *Dobývání znalostí z databází. (Knowledge Discovery in Databases),* Academia, Praha 2003.

[2] Good, P. I., Hardin, J. W.: *Common Errors in Statistics (And How to Avoid Them),* Wiley, 2006.

[3] Grünwald, P.: A Tutorial Introduction to the Minimum Description Length Principle. [online]. 2004, p. 80 [cit. 2014-07-15]. Available from: http://eprints.pascal-network.org/archive/00000164/01/mdlintro.pdf

[4] Hansen, M. H., Yu, B.: Minimum Description Length Model Selection Criteria for Generalized Linear Models. [online]. p. 20. Available from: http://www.stat.ucla.edu/ cocteau/papers/pdf/glmdl.pdf

[5] Huffman, D. A.: A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the I.R.E.* (September 1952), 10981102.

[6] Jiroušek, R.: Foundations of compositional model theory. *Int. J. General Systems* **40** (2011), 623–678.

[7] Jiroušek, R., Kratochvíl, V.: Foundations of compositional model theory: structural properties. *Int. J. General Systems* **44** (2015), 2–25.

[8] Jiroušek, R., Krejčová I.: Minimum Description Length Principle for Compositional Model Learning. In: *Integrated Uncertainty in Knowledge Modelling and Decision Making. Proceedings of IUKM 2015* (Van-Nam Huynh, Masahiro Inuiguchi, and Thierry Denoeux, eds.), LNAI, Vol.: 9376 (2015), 254–266.

[9] Kolmogorov, A.N.: Tri podchoda k opredeleniju ponjatija kolichestvo informacii. *Problemy Peredachi Informatsii*, **1** (1965), 3–11 [Engl. transl. accessible at http://alexander.shen.free.fr/library/Kolmogorov65_Three-Approaches-to-Information.pdf]

[10] Kullback, S., Leibler, R. A.: On information and sufficiency. *Annals of Mathematical Statistics* **22** (1951), 76–86.

[11] Lam, W., Bacchus, F. Learning Bayesian Belief Networks: An approach based on the MDL Principle. *Computational intelligence* **10** (1994), 269–293. [Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.127.5504&rep=rep1&type=pdf]

[12] Von Mises, R.: *Probability, statistics, and truth*. Courier Corporation, Mineola, New York, 1957 [Originaly published in German by Springer, 1928]

[13] Rényi, A.: In: *On measures of information and entropy. Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, (1960), 547–561.

[14] Ryan, T.: *Modern Regression Methods,* Wiley, 2009.

Optimisation in a Wholesale Company: A Supply Chain Design Problem

Petr Jirsak[1], Veronika Skocdopolova[2], Petr Kolar[3]

Abstract. The paper focuses on supply chain design problem, particularly on formulation and verification of a mathematical model. In our research, two interrelated mathematical models represent the supply chain design. The first one defines facility location problem and the second one elaborates vehicle routing problem. Supply chain design is one of the strategic challenges the companies deal with. The first model deals with the decision, where the company's warehouses should be placed. The results serve as an input for the second model based on vehicle routing problem suggesting the best routes from the warehouses to the customers.
The necessity of combining the opposing criteria leads to the challenge complexity. The quality of logistic services results in company competitive advantage. It requires higher density of logistic objects and number of routes. However, efficiency of supply chain is qualification if not decisive criterion for customers. It has to be incorporated in it.
The paper outlines the solution of the design challenge based on a case study of the wholesale company targeting wide range of its customers in the Czech Republic.

Keywords: vehicle routing problem, location set covering problem, supply chain design, cost, customer service.

JEL Classification: C61
AMS Classification: 90B06

1 Introduction

The supply chain design influences 80 % of supply chain costs [9]. The critical factor of any supply chain is the number and locations of facilities such as plants, suppliers, or warehouses. For the optimisation of the location and size of supply chain facilities, we use the supply chain network design. Our mathematical framework is the network modelling.

This paper elaborates a case study, in which a wholesale company at first wants to optimise number and location of its warehouses. The second task is to design the routes from the warehouses to customers. We formulated two interrelated models to solve these two optimisation tasks. The first one is based on a location set covering problem. The result of this model is a location of company's warehouses and an assignment of customers to the warehouses. The second model is a modification of a vehicle routing problem.

2 Theory and Methods

Many situations that occurs in the supply chain design can be presented on a network. The network is a directed connected graph without circuits [3]. The graph is a set of nodes and arcs. The nodes representing e.g. cities, customers or crossroads are connected with arcs (e.g. streets or roads). In the paper, we have utilised two problems defined in a graph: location set covering problem and vehicle routing problem.

2.1 Supply Chain Design

Supply chain design is a strategic discipline because of its relevance to company competitive advantage [1]. Logistic services have become critical source of competitiveness as the difference between functions and quality of products has vanished at the end of 20th century. Furthermore, supply chain disruption and sensitivity to supply chain performances require high concern of supply chain design. [2], [6]

[1] University of Economics, Prague, department of logistics, W. Churchill sq. 4, Prague, Czech Republic, petr.jirsak@vse.cz.
[2] University of Economics, Prague, department of econometrics, W. Churchill sq. 4, Prague, Czech Republic, veronika.skocdopolova@vse.cz.
[3] University of Economics, Prague, department of logistics, W. Churchill sq. 4, Prague, Czech Republic, petr.kolar@vse.cz.

Supply chain design creates set of nodes and appropriate arcs connecting them. The nodes can represent logistic facilities on different level of aggregation for instance a warehouse, warehouse zone or one working centre. Appropriate arcs connect the relevant nodes and for example take form of road, fork lift truck track or conveyer. Supply chain design defines infrastructure and processes that ensure production and delivery of required value added to customers [8]. Hence, customer service requirements, cost efficiency, technical bases, and ecology form particular supply chain network. The location set covering problem defines number and location of logistic objects. The recent papers focus on 42 factors but mainly on the following factors: demand level, cycle time, distance between nodes, demand volatility and delivery frequency [7]. Vehicle routing problem organizes effectiveness and efficiency in distribution of goods from logistic objects to the customers distributed in particular region.

2.2 The Location Set Covering Problem

The Location Set Covering Problem (LSCP) introduced in early 1970s was one of the first models developed to place emergency service facilities that incorporates a maximum service distance standard. The objective of the LSCP is to locate the smallest number of facilities that cover each demand node by one or more facilities. The demand node is covered, if there is a facility within a specified time or distance from the node. [4]

The objective of the LSCP can be modified to locate given number of facilities such the total distance (or time) from all the customers to assigned facility is minimal. The aim of the LSCP is to cover all the nodes (customers). The basic LSCP model [9] can be formulated as follows:

Minimize

$$z = \sum_i \sum_j dist_{ij} d_j Y_{ij} \tag{1}$$

Subject to

$$\sum_i Y_{ij} = 1, \quad \forall j \tag{2}$$

$$\sum_i X_i = P \tag{3}$$

$$Y_{ij} \le X_i, \quad \forall i, j \tag{4}$$

$$Y_{ij} \in \{0, 1\}, \quad \forall i, j$$

$$X_i \in \{0, 1\}, \quad \forall i$$

The objective function (1) minimises the total weighted distance from facilities (warehouses) to demand nodes (customers); parameter $dist_{ij}$ represents the distance between the warehouse i and the customer j; d_j is the demand of j-th customer. The constraint (2) assures that each customer j is assigned to exactly one warehouse, where Y_{ij} is a binary decision variable that equals 1, if the customer j is assigned to the warehouse i, and 0 otherwise. Exactly P facilities should be located in the area, where X_i is a binary decision variable that equals 1, if the facility is located in the location i, and 0 otherwise. The constraint (4) tie together the decision variables X_i and Y_{ij}.

In the terminology of supply chain management, the location set covering problem is known as a Facility Location Problem (FLP).

2.3 Vehicle Routing Problem

The Vehicle Routing Problem (VRP) is a modification of Travelling Salesman Problem (TSP). The objective of the TSP is to find the shortest circuit in the given graph that includes each node of the graph once only (the salesman has to visit all his customers and then go back home). Such a circuit is called Hamiltonian circuit. The VRP designs several routes from a single depot to a number of customers. The basic constraints of VRP are that each node (customer) is visited exactly once by exactly one vehicle and that each route starts and ends at the same depot. The objective is to design optimal delivery routes. The most common additional constraints of the VRP are linked to vehicle capacity restrictions, total time restrictions, time windows or precedence relations [9]. If there are different vehicles available with different capacity and cost, the heterogeneous fleet vehicle routing problem is considered. The basic VRP model is defined in [3].

3 Model Formulations

There are two main tasks in our supply chain design challenge. The first task is to locate company's warehouses and to assign the customers to the warehouses. The aim is to minimise the total cost that include transportation costs (transport from the warehouses to customers), storage costs (store rent) and personal costs (storekeeper wage). The second task is to design optimal delivery routes from the warehouses located in the first task to the assigned customers. To find the optimal solution, we formulated two interrelated models. The results of the first model serve as an input for the second model.

3.1 The Wholesale Company

The research subject is a SME (small and medium-sized enterprise) wholesale company employing less than 200 people. In the Czech Republic market, the company operates with annual turnover of 12 mil. EUR. It is a company trading painter´s goods targeting at wide range of customer segment of sole traders, small and medium drug stores and retail chain companies. There are 5000 stock keeping units in the company product portfolio. The company has 482 customers serviced by five wholesale warehouses. Each warehouse has fixed service area in the country. The current distribution structure is a result of past merges and acquisitions. Rented warehouses have predetermined distribution for a decade. However, once the company can redesign their distribution determining customer service and effectiveness, it will enhance its competitive advantage.

3.2 The Facility Location Problem

The first part of the supply chain design problem is to locate several warehouses in the Czech Republic while meeting specified requirements. We assume there is a warehouse in every city, in which one of the company's customer is located. The company does not plan construction of new warehouses, but it will rent a storage capacity in existing warehouses. To solve this part of the problem, we formulated the following model based on the location set covering problem:

Minimize

$$z = \sum_i \sum_j pricekm \cdot 2d_{ij} x_{ij} p_j + \sum_i \sum_j sc_i a_j x_{ij} + \sum_i \sum_j pc_i \frac{a_j}{w} x_{ij} \tag{5}$$

Subject to

$$\sum_i x_{ij} = 1, \quad \forall j \tag{6}$$

$$\sum_j x_{ij} \le ny_i, \quad \forall i \tag{7}$$

$$t_{ij} x_{ij} \le limit_j y_i, \quad \forall i, j \tag{8}$$

$$\sum_j a_j x_{ij} \ge minstock \cdot y_i, \quad \forall i \tag{9}$$

$$\sum_i y_i \le k, \tag{10}$$

$$x_{ij} \in \{0, 1\}, \quad \forall i, j$$

$$y_i \in \{0, 1\}, \quad \forall i$$

where the indexes i and j represent the company's customers (cities). There are two binary decision variables in the model; x_{ij} is equal 1 if the customer j is assigned to the warehouse in the city i, and 0 otherwise, and y_i is equal 1 if the warehouse is placed to the city i, and 0 otherwise. The meaning of the model parameters is as follows:

- *pricekm* – cost per 1 km;
- d_{ij} – distance in kilometres between customers i and j;
- p_j – importance of the j-th customer;

- sc_i – storage cost of 1 kg of goods per month;
- a_j – average monthly demand of j-th customer in kilograms;
- pc_i – personal cost per hour of the storekeeper in the i-th city;
- w – efficiency of a storekeeper (kg per hour);
- n – number of customers;
- t_{ij} – time distance in minutes between customers i and j;
- $limit_j$ – within this time in minutes the vehicle from a warehouse has to reach the customer j;
- *minstock* – minimal stock capacity in kilograms;
- k – maximal number of warehouses.

The objective function (5) minimises the total cost. It contains the transportation, storage and personal costs. The constraint (6) assures that each customer is assigned to exactly one of the warehouses. The second constraint (7) guarantees that customers are assigned only to cities, in which the warehouse is located. The requirements of the customer reach by within the given time limit is given by (8). The time limit can be set separately for each customer. The capacity of each warehouse has to be at least in a given limit (9). There should be maximum of k warehouses placed in the Czech Republic. If we omit the constraints (9) and (10), the optimal solution is to place a warehouse in every city.

The result of this model is a location of maximal k warehouses and an assignment of all the customers to these warehouses.

3.3 The Vehicle Routing Problem

Once the solution from the facility location problem model is retrieved, we can approach the second model. The model´s aim is to design the optimal delivery routes. The company wants to plan the delivery in two weeks (10 days). Based on the prerequisite, there are three types of customers: customers that order twice a week (*gr1*), once a week (*gr2*) and once in two weeks. For the delivery, the company uses two types of cars with different capacity and cost per km. Our formulated model employs four indexes – i and j represent the customers assigned to one warehouse ($i = j = 1$ represents the warehouse); k represents the type of the car; and l stands for the day (1…10). There are two binary decision variables in the model; x_{ij}^{kl} is equal 1, if the vehicle k goes from the customer i to the customer j on the day l, and 0 otherwise; y_i^l is equal 1, if there is a delivery to the customer i on the day l.

The model objective is to minimise the transportation costs:

$$z = \sum_i \sum_j \sum_k \sum_l pricekm_k d_{ij} x_{ij}^{kl} \tag{11}$$

where $pricekm_k$ is the cost per kilometre for the vehicle k, and d_{ij} is the distance in kilometres between customers i and j.

Each customer i (except the warehouse) has to get exactly pp_i deliveries:

$$\sum_j \sum_k \sum_l x_{ij}^{kl} = pp_i, \quad i \geq 2 \tag{12}$$

If the vehicle k on the day l enters the city of customer i, the same vehicle has to leave the city in the same day:

$$\sum_i x_{ij}^{kl} = \sum_i x_{ji}^{kl}, \quad \forall j, \forall k, \forall l \tag{13}$$

Each day l most $count_k$ vehicles k can leave the warehouse:

$$\sum_{j \geq 2} x_{ij}^{kl} \leq count_k, \quad i = 1, \forall k, \forall l \tag{14}$$

The following constraints assure the right construction of the routes by the capacity of the vehicles and by time requirements:

$$u_i^l + q_j - M(1 - x_{ij}^{kl}) \leq u_j^l, \quad \forall i, j \geq 2, \forall k, \forall l \tag{15}$$

$$at_i^l + st + t_{ij} - M(1 - x_{ij}^{kl}) \leq at_j^l, \quad \forall i, j \geq 2, \forall k, \forall l \tag{16}$$

$$u_i^l \le \sum_j \sum_k x_{ij}^{kl} \cdot cap_k, \quad \forall i, \forall l \tag{17}$$

$$u_i^l = 0, \quad i = 1, \forall l \tag{18}$$

$$x_{ij}^{kl} = 0, \quad \forall i, j = i, \forall k, \forall l \tag{19}$$

There are two artificial variables in these constraints; u_i^l represents the free capacity of the vehicle once it leaves the customer i on the day l; at_i^l is the arrival time to customer i in the day l. Parameter q_j stands for the demand of customer j for one delivery; M is a high value; st is the average service time; and t_{ij} is the time needed for the journey between customers i and j.

The total time (for which a vehicle is travelling) must not exceed 8 hours (480 minutes):

$$at_i^l + \sum_k t_{i1} x_{ij}^{kl} \le 480, \quad \forall i, \forall l \tag{20}$$

The following constraints assure the connection of decision variables:

$$\sum_j \sum_k x_{ij}^{kl} \le y_i^l, \quad i \ge 2, \forall l \tag{21}$$

$$\sum_{i \ge 2} \sum_l y_i^l = \sum_{i \ge 2} pp_i \tag{22}$$

For each day l, there should be at least one delivery:

$$\sum_i \sum_j \sum_k x_{ij}^{kl} \ge 1, \quad \forall l \tag{23}$$

The following constraints assure a regular schedule for those customers ordering twice a week (*gr1*). If there is a delivery on Monday (l = 1), the other delivery in the same week has to be on Wednesday (l = 3) or Thursday (l = 4). The constraints (25) – (28) are interpreted similarly:

$$y_i^1 \le y_i^3 + y_i^4, \quad i \in gr1 \tag{24}$$

$$y_i^2 \le y_i^4 + y_i^5, \quad i \in gr1 \tag{25}$$

$$y_i^3 \le y_i^1 + y_i^5, \quad i \in gr1 \tag{26}$$

$$y_i^4 \le y_i^1 + y_i^2, \quad i \in gr1 \tag{27}$$

$$y_i^5 \le y_i^2 + y_i^3, \quad i \in gr1 \tag{28}$$

These constraints assure that the schedule of deliveries is the same in both weeks for the customers ordering twice a week (*gr1*) or once a week (*gr2*):

$$y_i^l = y_i^{l+5}, \quad i \in gr1 \cup gr2, l \le 5 \tag{29}$$

All the decision variables are binary, the artificial variables are nonnegative:

$$x_{ij}^{kl} \in \{0, 1\}, \quad \forall i, \forall j, \forall k, \forall l, \; y_i^l \in \{0, 1\}, \quad \forall i, \forall l$$

$$u_i^l \ge 0, \quad \forall i, \forall l, \; at_i^l \ge 0, \quad \forall i, \forall l$$

The described model has to be solved for each warehouse repeatedly. The result of the model is a schedule of delivery routes.

4 Results and Conclusion

To solve both models, we used the Gurobi 7.0.1 solver via MPL 5.0 modelling system connected with MS Excel for importing the input data and exporting the results.

The company has 482 customers. When we solve the first model without the constraints (9) and (10), the result is a warehouse in each city. If we omit only the constraint (10), the optimal solution is to place 14 warehouses for the minimal stock capacity of 45 tons of goods. We try to solve the model with different values of k. The lower the k is the higher the total cost is. For $k = 3$ there is no feasible solution.

For the second model, we chose the results with four warehouses. We prepared data for solving four different VRPs. A problem occurs during solving the VRPs. After many hours of solving (12 hours and more), the solver reached a feasible solution but not an optimal solution. The gap between the objective value of the best feasible solution and the best bound of the objective value was about 50 %. The problem is in the size of the problems – each VRP deals with more than 100 customers, which means more than 200,000 binary variables.

For further research, we shall simplify the VRP model. It is possible to solve the model only for the biggest customers (*gr1*) and add the other customers to their schedule later on. Another possibility is to strengthen the computational capacity or to try to find better feasible solution via heuristic methods, such as [5].

Acknowledgements

The presented research is supported by the grant No. IP305026 of the Internal Grant Agency of the Faculty of Business Administration, University of Economics, Prague.

References

[1] Christopher, M.: *Logistics and Supply Chain Management*. FT/Pearson Education, 2011.
[2] Christopher, M. and Peck, H. Building the Resilient Supply Chain. *International Journal of Logistics management* **15** (2004) 1–13.
[3] Eiselt, H. A, and Sandblom, C.-L.: *Integer Programming and Network Models.* Springer-Verlag, Berlin, 2000
[4] Eiselt, H. A, and Sandblom, C.-L.: *Operations Research: A Model-Based Approach.* Springer-Verlag, Berlin, 2012.
[5] Fabry, J., Korenar, V., and Kobzareva, M.: Column Generation Method for the Vehicle Routing Problem – The Case Study. In: *Mathematical Methods in Economics 2011*. Professional Publishing, Prague, 2011, 140–144.
[6] Lambiase, A., Mastrocinque, E., Miranda, S., and Lambiase, A.: Strategic Planning and Design of Supply Chains: A Literature Review. *International Journal of Engineering Business Management* **5** (2013), 5–49.
[7] Mangiaracina, R., Song, G., and Perego, A.: Distribution network design: a literature review and research agenda. *International Journal of Physical Distribution and Logistics* **45** (2015), 506–531.
[8] Sapiro, J. F. *Modeling the Supply Chain*. Thomson, 2007.
[9] Watson, M., Lewis, S., Cacioppi, P., and Jayaraman, J.: *Supply Chain Network Design: Applying Optimization and Analytics to the Global Supply Chain*. Pearson, New Jersey, USA, 2014.

Investment projects threshold value simulation

Petr Jiříček[1], Stanislava Dvořáková[2]

Abstract. The aim of the paper is determining the threshold values of an investment project representing so called cash break-even points for its individual stages. Threshold value can be defined as such amount of cash-flow that induces the break-even value of investment project efficiency. The threshold values will be simulated for both conventional and nonconventional projects (where oscillating between positive and negative cash-flows occurs) and depending on discount rate. The simulations will also show tests of threshold values for critical cash-flow amounts (stress test) in pre-defined stages of the investment project.

Keywords: Threshold value, investment projects, cash break-even point, discount rate.

JEL Classification: C20, H43
AMS Classification: 65H04

1 Introduction

In economic theory and practice, the break-even point analysis is conceived as a significant part of investment project efficiency evaluation. In this context, a beneficial impact of an investment in the operational stage means that sufficient means for investment costs payment are generated and an overall positive effect occurs after reaching the break-even point. The classical break-even point analysis in the accounting conception provides information on the value of a variable determined in an accrual way (income, expenses). It means that we are able to determine such amount of production at which the product or service sales revenues cover production costs.

If we evaluate the investment efficiency on the basis of cash-flow, we accept the definition of the so called cash break-even point. In this concept, the break-even point is the moment when cash income from the investment (inflows) equals the expenditures related to the investment (outflows).

The modern approach to investment efficiency evaluation uses discounted cash-flows, i.e. it considers a different time value of money in individual stages of the investment process. The most often used method for evaluating the economic efficiency of investments is the Net Present Value method [7]. In this concept, cash-even point can be defined as the so called financial break-even point, i.e. the moment when the Net Present Value (NPV) of the evaluated investment equals zero. It uses break-even chart showing the present values of cash-inflows and outflows under different assumptions about unit production Q [2].

In case of public projects we consider generating cash-flow also on the non-commercial basis, i.e. we think in the categories of social utility. The total social utility of a public project, e.g. in projects financed from the European funds, is defined as Economic Net Present Value [5]. In this conception cash-even point can be defined as so called social break-even point when public project benefits that contain financial cash-flows including cash-flows from externalities balance the costs issuing from the public project. It is the point when Economic Net Present Value (ENPV) of a public project financed from the European funds equals zero. Here the optimum level occurs when marginal benefits (B) equal marginal costs (C) related to a unit of production Q, i.e. it holds that dB/dQ = dC/dQ [1].

Another approach to determining social break-even point while meeting the requirement $ENPV = 0$ is offered by the definition of *threshold value* cash-flow [4]. It is determining such marginal value of cash-flow in individual years of the investment cycle that causes an indifferent value of social utility from the public project implementation.

2 Objective and methods

The presented paper aims at simulating social break-even point by means of modelling the *threshold value* cash-flow in individual stages of the investment cycle of a real public project. It is a case study of an investment project within building the system of e-Governance in the Olomouc Region, subsidized from the European funds [6]. To

[1] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, jiricek@vspj.cz.
[2] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 586 01 Jihlava, Czech Republic, stanislava.dvorakova@vspj.cz.

define the model, we will use the European Union methodology for evaluation of the social utility of investment projects financed from the European funds for the 2014–2020 cohesion period [5]. Model variants for the calculation of *threshold value* cash flow will be used on the grounds of so-called economic cash flows including both purely commercial, so-called financial cash flow and externalities related to the project and quantified into the value of the cash flows.

The *threshold value* simulation will follow up on the conventional project scenario [4]. For the presented paper, three further defined crisis scenarios will be modelled besides the current project scenario of the Olomouc Region. The scenarios will be set at the level of stress tests, which can be considered a sensitivity analysis method [8], when we make a change of the critical input variable on a large scale.

A. Original conventional project scenario

B. Conventional project; a crisis situation occurs in the building stage of the investment cycle due to postponing the investment building by one year and consequently there is a sharp change of positive cash flow to negative cash flow (the third year of the cycle = 2013).

C. Nonconventional project; a crisis situation occurs in the operational stage of the investment cycle due to the necessity to make additional investments and consequently there is a sharp change of positive cash flow to negative cash flow (the seventh year of the cycle = 2017).

D. Nonconventional project; a crisis situation occurs in the liquidation stage of the investment cycle due to a loss of revenues resulting from a technological innovation of other service providers and consequently there is a sharp change of positive cash flow to negative cash flow (the 12th year of the cycle = 2022).

Results and discussions

The Net Present Value method in case of public projects of developmental nature ($ENPV$) states what cash flow caused by the investment is left after deducting investment costs in pre-projected lifetime.

$$ENPV = \sum_{t=0}^{n} \frac{CF_t}{(1+k)^t}, \tag{1}$$

where CF_t is the cash flow of the public development project (i.e. the benefits) minus costs, losses of the project,

k is the common social discount rate (the required return on public investment from the perspective of the investor or donor of the project),

n is the period of economic lifespan of the project (in our model case $n = 12$),

$ENPV$ is economic net present value of the public project reflecting the benefit from the project from a wider social economic view of the given investment.

This relation can be defined as the project investment curve [3]. In the following Figure 1 the $ENPV$ curves for the above mentioned scenarios A – D are displayed.

The criteria for evaluating the efficiency of a public developmental project are stated by methodology of the European Union for evaluating the social utility from projects financed from the European [5]:

$ENPV > 0$ – The investment yields positive social utility

$ENPV < 0$ – The investment yields negative social utility

$ENPV = 0$ – The investment yields zero social utility (the net economic welfare does not change).

Now we will examine the threshold values of input variables, i.e. cash flows of the project for the selected discount rate. It will be the social discount rate determined by the project donor, i.e. the European Union, for financial support from the European funds in the programming period of 2014–2020. The rate of 5% p.a. is valid for financing public projects of the EU member states from structural funds and Cohesion fund.

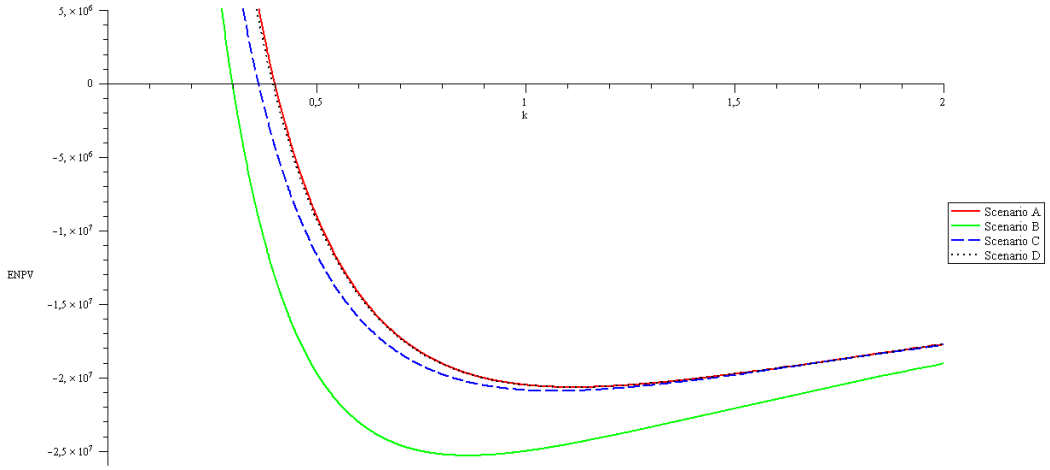


Figure 1 Investment curves of the project $ENPV = f(x)$ for individual scenarios

From formula (1) we can draw the general formula for calculation of threshold value in individual years of the investment cycle, i.e. the variable CF_i (for $i = 0, \dots, n$)

$$CF_i = ENPV \cdot (1+k)^i - \sum_{t=0}^{i-1} \frac{CF_t}{(1+k)^{t-i}} - \sum_{t=i+1}^{n} \frac{CF_t}{(1+k)^{t-i}} = -\sum_{t=0}^{i-1} \frac{CF_t}{1.05^{t-i}} - \sum_{t=i+1}^{n} \frac{CF_t}{1.05^{t-i}}, \tag{2}$$

while in our case we consider the condition $ENPV = 0$ and the discount rate $k = 0.05$.

The model for calculation of *threshold value* will now be tested on a real proposed project of e-Governance, financed from the European funds, which was proposed to be implemented in years 2010–2022. In Tab. 1 and Tab. 2 we will find out the threshold values of critical variables of projects, i.e. such values of cash flows in individual years of model project scenarios when the economic net present value of the project equals zero and the project does not provide social utility. The *threshold value* thus shows the critical value of cash flows in individual years of the project lifetime for given scenarios.

		Scenario A (original proposal)		**Scenario B** (building phase)	
t	**Year**	**Project CF in mil. CZK**	**Threshold value in mil. CZK**	**Project CF in mil. CZK**	**Threshold value in mil. CZK**
0	2010	–0.58	–166.45	–0.58	–135.31
1	2011	–57.27	–231.44	–57.27	–198.75
2	2012	–0.30	–183.18	–0.30	–148.85
3	2013	35.58	–156.44	–0.46	–156.44
4	2014	35.58	–166.04	35.58	–128.19
5	2015	35.58	–176.12	35.58	–136.38
6	2016	35.58	–186.71	35.58	–144.98
7	2017	35.58	–197.82	35.58	–154.01
8	2018	15.58	–229.49	15.58	–183.49
9	2019	29.58	–227.74	29.58	–179.44
10	2020	29.58	–240.61	29.58	–189.89
11	2021	29.58	–254.12	29.58	–200.87
12	2022	29.58	–268.31	29.58	–212.39

Table 1 Threshold value for conventional scenarios of the investment project of e-Governance – scenarios A, B

In the following Figures 2 to 5 we can see graphic presentation of project cash flows in individual simulated scenarios of e-Governance investment project and the *threshold value* of cash flows when loss of social utility occurs at the critical value of $ENPV = 0$.

		Scenario C (crisis in operational phase)		**Scenario D** (crisis in liquidation phase)	
t	**Year**	**Project CF in mil. CZK**	**Threshold value in mil. CZK**	**Project CF in mil. CZK**	**Threshold value in mil. CZK**
0	2010	–0.58	–134.47	–0.58	–149.00
1	2011	–57.27	–197.86	–57.27	–213.11
2	2012	–0.30	–147.92	–0.30	–163.93
3	2013	35.58	–119.42	35.58	–136.23
4	2014	35.58	–127.17	35.58	–144.82
5	2015	35.58	–135.30	35.58	–153.84
6	2016	35.58	–143.85	35.58	–163.32
7	2017	–9.42	–197.82	35.58	–173.26
8	2018	15.58	–182.24	15.58	–203.70
9	2019	29.58	–178.13	29.58	–200.67
10	2020	29.58	–188.52	29.58	–212.18
11	2021	29.58	–199.42	29.58	–224.27
12	2022	29.58	–210.87	–1.76	–268.31

Table 2 Threshold value for nonconventional scenarios of the investment project e-Governance – scenarios C, D

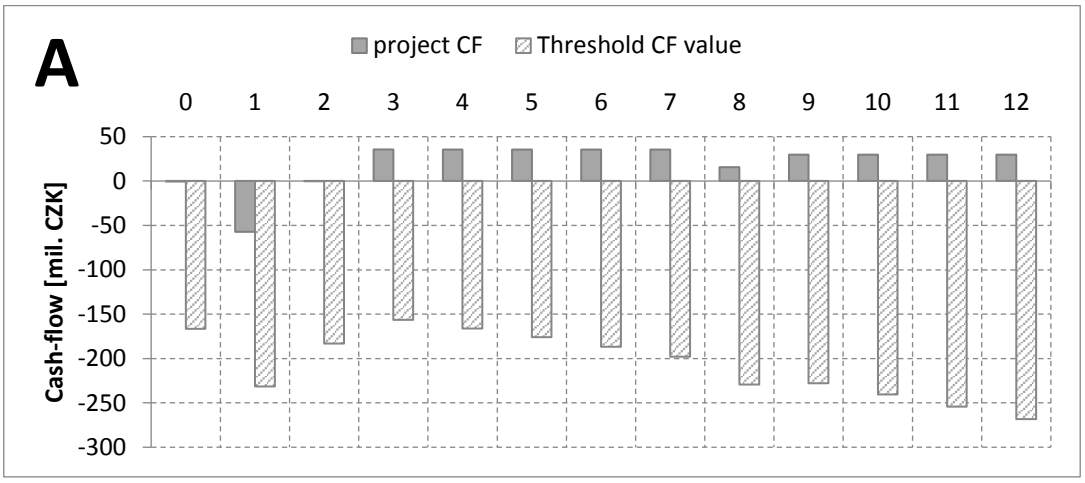


Figure 2 Cash flow threshold values of conventional project – scenario A

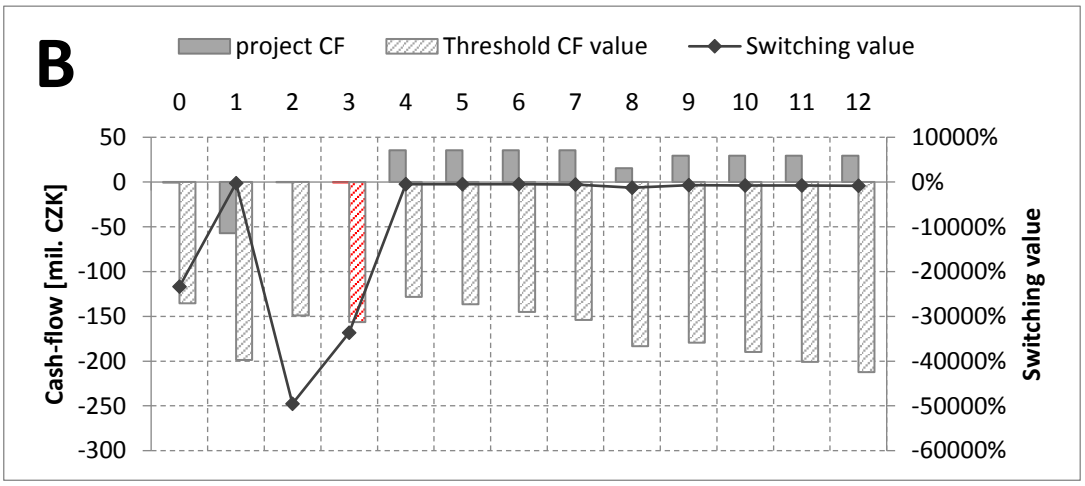


Figure 3 Cash flow threshold values of conventional project – scenario B

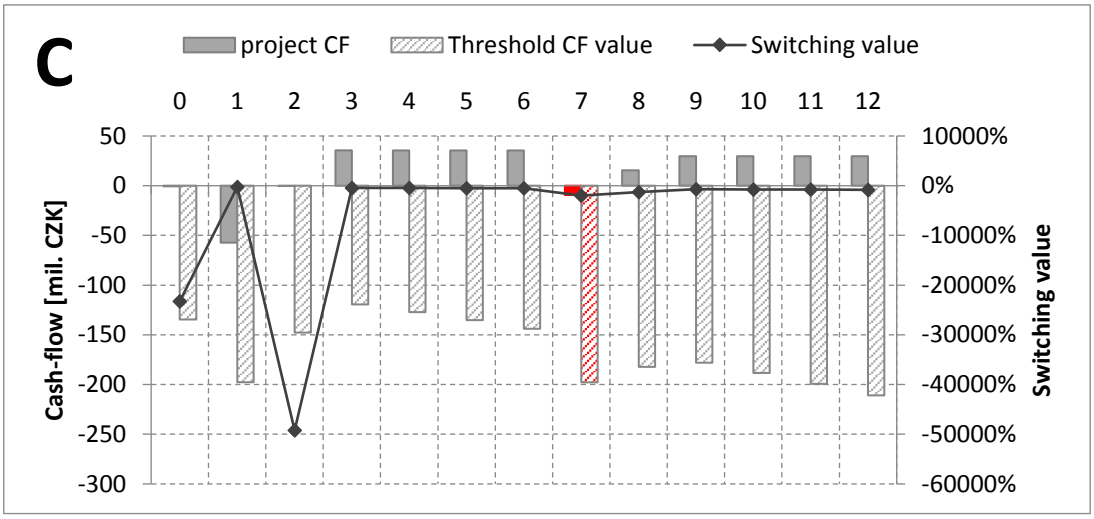


Figure 4 Cash flow threshold values of nonconventional project – scenario C

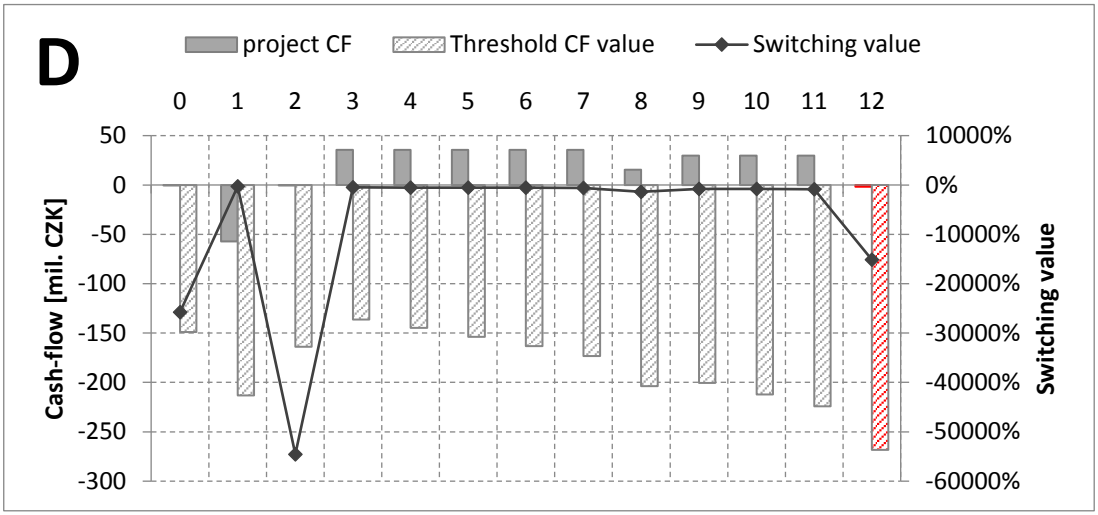


Figure 5 Cash flow threshold values of nonconventional project – scenario D

Based on the results in Tab. 1–2 we can state that the threshold value in examined model scenarios (proposed conventional project, conventional project with a crisis in building stage, nonconventional project with a crisis in operational stage, and nonconventional project with a crisis in liquidation stage) is gradually growing from the third year of the project cycle. An exception from the trend is the years with simulation of critical cash flow value for individual crisis scenarios, i.e. for scenario B it is the third year, for scenario C it is the seventh year and for scenario D it is the twelfth year. We can also see (Fig. 1–2) that in all scenarios the trend of growing negative threshold value is broken in the eight year of the project cycle.

3 Conclusion

Determining the threshold value of a public investment project can be defined as such a value of cash flow in individual years of the project cycle when the social break-even point is reached. It is the situation when zero social utility of the project is reached as measured by economic net present value. The threshold value determines the sensitivity of the output variable of the project to the change of the input variable. Low threshold values correspond to high sensitivity and high threshold values on the contrary mean low sensitivity of social utility of the project to the change of cash flow. Based on calculations presented in Tables 1 and 2, we can say that the project of e-Governance has low sensitivity of social utility to changes of cash flows in individual years of the project. The presented model offers possible directions of further progress: either for finding causes of low sensitivity of the project by further analysis of the structure of determined critical cash flows in the project cycle (to individual

revenues and expenditures) or for analyzing the causes of deviations in the trend of threshold value – here e.g. in the mentioned eighth year of the project cycle.

Acknowledgements

Supported by the grant No. IGS č. 1170/4/179 of the College of Polytechnics Jihlava.

References

[1] Boardman, A. and Greenberg, D. and Vining, A. and Weimer, D.: *Cost-Benefit Analysis. Concepts and Practise.* 4th edition. Harlow: Pearson Education Limited, 2014. 500 p. ISBN 978-1-29202-191-1

[2] Brealey, R. and Myers, S. and Allen, F.: *Principles of Corporate Finance*. 2th. edition. New York: McGraw-Hill Irwin, 2011, 524 s. ISBN 978-0-07-353074-1

[3] Dvořáková, S. and Jiříček, P.: *Modelling Financial Flows of Development Projects Subsidized from European Funds*. In: 31st International Conference on Mathematical Methods in Economics. Jihlava: VŠPJ. 2013. p. 119-128. ISBN 978-8097035-79-4

[4] Dvořáková, S. and Jiříček, P.: *Switching value of public projects.* In 33rd International Conference on Mathematical Methods in Economics. x. 1 CD ROM. Plzeň: University of West Bohemia. 2015. s. 161-166. ISBN 978-80-261-0539-8

[5] European Commission: *Guide to Cost-Benefit Analysis of Investment Project*. Brusel: DG Regional and Urban Policy. 2015. ISBN 978-92-79-34796-2

[6] Olomoucký kraj. *Studie proveditelnosti projektu Technologické centrum a elektronické spisové služby* [online]. 2011.[cit. 2017-04-18]. http://www.kr-olomoucky.cz/clanky/dokumenty/202/priloha-c1-studie-proveditelnosti-technologicke-centrum-a-elektronicka-spisova-sluzba.pdf

[7] Osborne, M.: *A resolution to the NPV-IRR debate?* The Quarterly Review of Economics and Finance 50(2), 2010, 234-239

[8] United States Environmental Protection Agency. *Guidance on the Development, Evaluation, and Application of Environmental Models* 2009. [online]. Washington: EPA/100/K-09/003. 2009 [cit. 2017-12-03]. www.epa.gov/crem

Robust regression estimators: A comparison of prediction performance

Jan Kalina[1], Barbora Peštová[2]

Abstract. Regression represents an important methodology for solving numerous tasks of applied econometrics. This paper is devoted to robust estimators of parameters of a linear regression model, which are preferable whenever the data contain or are believed to contain outlying measurements (outliers). While various robust regression estimators are nowadays available in standard statistical packages, the question remains how to choose the most suitable regression method for a particular data set. This paper aims at comparing various regression methods on various data sets. First, the prediction performance of common robust regression estimators are compared on a set of 24 real data sets from public repositories. Further, the results are used as input for a metalearning study over 9 selected features of individual data sets. On the whole, the least trimmed squares turns out to be superior to the least squares or M-estimators in the majority of the data sets, while the process of metalearning does not succeed in a very reliable prediction of the most suitable estimator for a given data set.

Keywords: robust estimation, linear regression, prediction, outliers, metalearning

JEL classification: C14
AMS classification: 62G35

1 Robust regression

Regression represents an important methodology for solving numerous tasks of applied econometrics. Throughout this paper, the standard linear regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + e_i, \quad i = 1, \ldots, n,$$

is considered with a continuous response Y explained by the total number of p regressors (independent variables) under the presence of random errors $e_1, \ldots, e_n$. The least squares estimator, which is the standard estimation tool for the linear regression model, is however well known to be too vulnerable to the presence of outlying measurements (outliers).

Within the framework of robust statistic, a variety of robust estimators was proposed as an alternative to the least squares [4, 9]. These tools are more suitable estimation techniques for data contaminated by outliers and thus represent important tools applicable to econometrics [5]. On the whole, M-estimators are the most commonly used robust methods. However, they do not possess a high breakdown point which has become of one fundamental robustness measures [2]. Thus, we may open a question if the least trimmed squares (LTS) estimator of [11] with a high breakdown point should not become more widely used in (not only) econometric applications.

Nevertheless, a systematic comparisons of the performance of robust regression estimator is missing and there remains a lack of comparisons of robust estimators on real data. Indeed, insufficient comparisons can be described as one of main reasons why robust methods have not much penetrated to real applications [6].

This paper has the following structure. Section 2 presents a study comparing the prediction performance of several common robust regression estimators over 24 real data sets. In Section 3, it is attempted to learn knowledge from the results of Section 2 by means of metalearning, i.e. learning a classification rule predicting the best estimator for a given data set based on its selected features. Finally, Section 4 concludes the paper and discusses also limitations of the presented study.

2 A direct comparison of robust regression methods

In this section, a study comparing various robust regression estimators on a database of real data sets is presented together with the results. First, we carefully selected 24 publicly available data sets. These are listed in Table 1, where also their sources are cited. The data sets, which are chosen, contain a continuous response and one or more continuous regressors allowing to consider a meaningful and interpretable linear regression model. As their titles indicate, a large portion of the data sets comes from the economics domain. We did not include data sets if no

[1]Institute of Computer Science CAS, Pod Vodárenskou věží 2, 182 07 Praha 8, Czech Republic, kalina@cs.cas.cz
[2]Institute of Computer Science CAS, Pod Vodárenskou věží 2, 182 07 Praha 8, Czech Republic, pestova@cs.cas.cz

information was available concerning a prior cleaning, pre-processing and quality control of the data. We did not include data sets with high multicollinearity. If a missing value appears in the data set, then the whole observation (all variables corresponding to the particular measurement) is ignored. The number of variables in the data sets ranges from 1 to 6 and the number of observations between 11 and 8088.

In the 24 selected data sets, only suitable regressors are considered and observations with missing values are omitted. Five important examples of linear regression estimators are used, namely the least squares, M-estimators with Huber's or Hampel's score function [4] and the least trimmed squares (LTS) [11] with trimming constant $h = \lfloor 0.5n \rfloor$ and $h = \lfloor 0.75n \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$. All computations of this paper were programmed in R software package, exploiting packages class, e1071, lmtest, MASS, moments, robustbase and robustreg.

To measure the prediction error of each estimator, the mean square prediction error $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/n$ is used, where $\hat{Y}_i$ denotes the fitted value (i.e. estimator of Y_i) by means of the particular regression estimator. This prediction error is evaluated in an autovalidation and a leave-one-out cross validation study. First, an autovalidation is considered where the regression estimator is computed over all n observations and the prediction is applied to the same n training observations. A standard leave-one-out cross validation is considered a more reliable approach and an attempt to evaluate a prediction performance over independent data. The computed values of the prediction performance of the 5 estimators were replaced by ranks. If ties occur, i.e. two estimators have the same prediction performance, then they obtain the same value of the rank. Table 2 presents the results in this form of ranks.

The results of the autovalidation give a false impression that the least squares estimator is superior to all remaining methods. It is remarkable how the more reliable results of the leave-one-out study are very different from the autovalidation. The LTS estimator is a clear winner of the study if a suitable h is chosen. The dependence on the choice of h seems rather strong. Although it is difficult to interpret the results without additional analyses, which will be done below, the performance of M-estimators is rather similar to that of the least squares, not only for data without outliers. Sometimes it even happens that Hampel's M-estimator coincides with the least squares, presumably for non-contaminated data sets with residuals with a distribution very close to normal.

These comparisons of the prediction performance do not however answer the question which factors are responsible for determining which method is the best for a particular data set. To extract such knowledge over the 24 data sets, we perform metalearning exploiting the results of Section 2.

3 Comparison of estimators by means of metalearning

The aim of this section is to learn a classification rule from the results of Section 2, which would allow to predict the most suitable regression estimator for a particular given data set. To do this, a classification analysis called metalearning will be performed over the 24 data sets and their selected features.

Metalearning is a general approach to learning about suitability of particular methods or algorithms over given (training) data sets. It is one of important data mining directions aimed at learning a classification rule, i.e. performing a supervised learning. Results over the training data sets play the role of a prior knowledge for other (new) data sets, similarly to Bayesian reasoning [3], and can be subsequently applied for selecting a suitable method or predict the performance of a given method.

Metalearning has been repeatedly used as a practical tool helpful with searching for the most suitable method for a particular real data set, mainly in the context of classification rules, association rules, decision trees [10] or algorithms for operation research [12] but very rarely to various (mostly nonparametric) methods of regression modelling [1].

Metalearning requires to select an appropriate and possibly small set of features which are computed for all training data sets. The features can be considered metadata, i.e. the metalearning does not use the original data sets but only the features. There are recommendations for choosing the features to include the size of the data set, signal-to-noise ratio, or violations of statistical assumptions. In our task, such features should be selected which are believed to relevant for evaluating the prediction performance of robust regression estimators. Thus, we selected the following set of 9 features:

1. The number of observations n,
2. The number of variables p,
3. The ratio n/p,
4. Normality of residuals, evaluated as the p-value of the Shapiro-Wilk test,
5. Skewness,
6. Kurtosis,
7. Coefficient of determination R^2,
8. Percentage of outliers estimated by the LTS in the form $\frac{1}{n}\sum_{i=1}^{n} I[u_i/\hat{\sigma} > 2.5]$ suggested by [11], where

Data set			Autovalidation					Leave-one-out					
Index	Name	Source	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(*)
1	Aircraft	[9]	1	3	2	5	4	5	3	4	1	2	4
2	Ammonia	[13]	1	3	2	5	4	5	3	4	1	2	4
3	Auto MPG	[8]	1	3	2	5	4	5	3	4	2	1	5
4	Cirrhosis	[13]	2	3	1	4	5	2.5	1	2.5	5	4	2
5	Coleman	[9]	1	2	4	5	3	1	2	4	5	3	1
6	Delivery	[9]	1	2	5	4	3	5	4	2	3	1	5
7	Education	[9]	1	3	2	5	4	5	1	3	4	2	2
8	Electricity	[13]	1	3	2	5	4	2	3	1	5	4	3
9	Employment	[13]	1	3	2	4	5	5	3	4	1	2	4
10	Furniture 1	[7]	1.5	3	1.5	4	5	2	3	1	4	5	3
11	Furniture 2	[7]	1	2	4	5	3	2	4	3	1	5	4
12	GDP growth	[7]	1.5	3	1.5	4	5	5	2	1	4	3	3
13	Houseprices	[13]	2	3	1	5	4	4.5	2	4.5	3	1	5
14	Housing	[8]	1	2	3	5	4	3	1	2	5	4	2
15	Imports	[13]	1	3	2	5	4	5	3	4	1	2	4
16	Kootenay	[9]	1.5	3	1.5	5	4	1.5	3	1.5	4	5	1
17	Livestock	[13]	1	3	2	4	5	3	5	4	1	2	4
18	Machine	[8]	1	2	3	5	4	5	4	3	1	2	4
19	Murders	[13]	1	3	2	5	4	5	3	2	1	4	4
20	NOx	[9]	1	3	2	5	4	4	2	3	5	1	5
21	Octane	[13]	1	3	2	4	5	2	3	1	4	5	3
22	Pasture	[13]	1	3	2	5	4	5	3	4	1	2	4
23	Pension	[9]	2	3	1	5	4	2	4	3	1	5	4
24	Petrol	[13]	1	3	2	5	4	5	2	3	4	1	5

Table 1 The list of 24 investigated data sets together with ranks corresponding to the mean prediction errors of five regression estimators for each of the data set. The methods evaluated in an autovalidation and a leave-one-out cross validation study include: (1) least squares, (2) Huber's M-estimator, (3) Hampel's M-estimator, (4) LTS with $h = \lfloor 0.5n \rfloor$ and (5) LTS with $h = \lfloor 0.75n \rfloor$. The column (*) contains the index of the best method according to the leave-one-study; this factor variable serves as a response for the subsequent classification task of metalearning.

$u_1, \ldots, u_n$ are residuals obtained by the LTS with $h = 0.5$, σ^2 is the common variance of $e_1, \ldots, e_n$ and $\hat{\sigma}$ is the estimate of σ obtained by the LTS with $h = 0.5$,

9. Heteroscedasticity evaluated as the p-value of the Breusch-Pagan test, which considers an auxiliary model for the errors in the form

$$\mathsf{var}\, e_i = \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_p X_{pi} + \gamma_i, \quad i = 1, \ldots, n$$

with parameters $(\alpha_0, \alpha_1, \ldots, \alpha_p)^T$ and random errors $\gamma_1, \ldots, \gamma_n$.

Data	Feature								
set	1	2	3	4	5	6	7	8	9
1	23	4	0.17	0.69	0.21	3.02	0.88	0.04	0.07
2	21	3	0.14	0.82	-0.19	3.11	0.91	0	0.18
3	392	4	0.01	0	0.71	4.05	0.71	0.03	0
4	46	4	0.09	0.11	-0.21	2.07	0.81	0	0.61
5	20	5	0.25	0.15	0.51	5.09	0.91	0.05	0.33
6	25	2	0.08	0.27	0.03	3.07	0.96	0.04	0.00
7	50	3	0.06	0.93	0.26	2.71	0.59	0.02	0.00
8	16	3	0.19	0.22	0.78	3.84	0.92	0.06	0.13
9	16	6	0.38	0.48	0.42	2.44	1.00	0	0.87
10	11	1	0.09	0.22	0.63	2.11	0.99	0.00	0.92
11	11	1	0.09	0.002	-1.89	5.82	0.49	0.09	0.10
12	22	1	0.05	0.70	-0.19	2.09	0.92	0.05	0.38
13	28	5	0.18	0.26	0.20	2.22	0.93	0.04	0.47
14	506	5	0.01	0	5.45	46.8	0.27	0.03	0
15	18	3	0.17	0.38	0.27	2.12	0.97	0.11	0.03
16	13	1	0.08	0.88	0.04	2.27	0.00	0.08	0.29
17	19	4	0.21	0.77	0.42	2.76	0.94	0.05	0.54
18	209	6	0.03	0	1.50	14.0	0.86	0.03	0
19	20	3	0.15	0.14	0.68	3.18	0.82	0	0.34
20	8088	3	0	0	-0.12	3.18	0.66	0.01	0
21	82	4	0.05	0.72	0.12	2.54	0.91	0.01	0.58
22	67	3	0.04	0.44	0.45	3.90	0.86	0.04	0.01
23	18	1	0.06	0.94	0.004	2.23	0.84	0	0.41
24	48	4	0.08	0.02	1.06	5.30	0.68	0.06	0.16

Table 2 Values of 9 selected features (characteristics of individual data sets) defined in Section 3 evaluated for each of the 24 data sets.

The aim of the metalearning is to construct a classification rule based on the 24 training data sets, where 9 selected features serve as independent variables and the response variable taken as the last column of Table 1. This is a simple classification task to 5 groups, where each data set plays the role of a single observation, and we solve it by several available classifiers, including linear discriminant analysis (LDA), a linear support vector machine (SVM), and a k-nearest neighbor (k-NN). Here, the last method is the most comprehensible one assigning a new observation according to the k-nearest neighbors in terms of the Euclidean distance, which is the most common measure in metalearning applications.

First, we evaluated the classification performance in an autovalidation study, in which the training 24 data sets are classified to one of the 5 groups. Such approach is however known to be severely biased in various applications. Therefore, a leave-one-out cross validation study is also performed as a preferable attempt for an independent validation.

The metalearning results over 24 data sets are presented in Table 3, where the ratio of correctly classified cases is presented. We consider the autovalidation results misleading, for example the k-nearest neighbor classifiers (k-NN) with $k = 1$ represents a tautology. The results of the leave-one-out cross validation study are very different from autovalidation results. While SVM performs superior to other classifiers, we must say that all the methods perform poorly in predicting the best regression method. In fact, no classification rule is able to find the best

regression estimator more often than in $40\ \%$ of cases. This does not however represent a worse result compared to those in other metalearning studies [1].

Method	Autovalidation	Leave-one-out
LDA	0.67	0.29
SVM (linear)	0.71	0.38
k-NN (k=1)	1.00	0.29
k-NN (k=3)	0.58	0.29
k-NN (k=5)	0.54	0.33

Table 3 Correct classification performance in the metalearning task over the 24 data sets. The classification rule for the best regression estimator is learned not over the original data sets, but using only values from Table 2 and the indicator of the best method obtained from Table 1. It is evaluated in an autovalidation and a leave-one-out cross validation study.

4 Conclusions

The aim of this paper is to investigate the prediction performance of robust regression estimators and to compare them on real data sets. For this task, we selected 24 real data sets with different properties, while some of them come from the economic domain.

First, a direct comparison of the prediction performance of individual estimators in Section 2 shows that there is no single method uniformly better than all remaining ones. It is instructive to see the big difference between results of the autovalidation and leave-k-out studies. A further leave-one-out study however reveals that the LTS performs in a much more reliable way compared to the least squares or M-estimators in the majority of data sets. If a given data set is not contaminated by outliers, then the LTS often loses its prediction performance rather slightly compared to other estimators and can be also recommendable. If however a data set is contaminated, then the least squares and M-estimators reduce their performance rather drastically. Additional computations also show that the results of the leave-k-out cross validation strongly depend on the value of k.

Statistical estimators are commonly compared by means of theoretical investigations, which would be however too complicated (or perhaps infeasible) for comparisons of robust regression estimators under various assumptions. Another problem of direct comparisons is a very complicated or unpredictable interpretation of the results. Therefore, after a direct comparison of regression estimators by means of computations on real data sets, this paper resorts to a metalearning study described in Section 3. Our exploiting the metalearning principles is rather similar to the approaches of various references, which are however most devoted to heuristic tools of computer science and not to statistical methods. We also verified the metalearning to be computationally well feasible even on a standard PC.

Metalearning allows to exploit knowledge from previously analyzed data sets for a given (new) data set. In the same way, metalearning may bring interpretability and comprehensibility to any large numerical simulations which are common in statistics or econometrics. While the mean prediction error is used, its robust alternative in the form of a trimmed mean prediction error reveals in additional computations the LTS estimator to be even more successful. The superiority of the LTS may be conjectured but only in terms of prediction, while severe disadvantages of the estimator (low efficiency or local sensitivity) are not revealed. In addition, there are rather small differences between least squares and M-estimators.

The metalearning part of our computations can be understood as a pilot study aimed at more intensive comparisons of various regression methods, e.g. for finding the best weights for the least weighted estimator [6]. The computations allowed us to find out practical problems encountered in metalearning. Although metalearning has been described as a trustworthy tool in computer science literature, our experience summarized in the following drawback and limitations of metalearning casts serious doubts on the process of metalearning itself and we see its potential for extracting knowledge from very different data sets (i.e. from different domains) as limited and controversial. Practical limitations or disadvantages, not seriously discussed in literature, include:

- The results are heavily influenced by the choice of the data sets, which can be hardly selected in a representative way. In addition, it is impossible to consider a uniform distribution over all data sets, which would be demanded for the metalearning process.
- The choice of features is crucial. Various relevant features may be however forgotten, not available, or not formalized (e.g. prior data normalization, prior feature selection, parameter tuning). Highly variable features may lead to a bias, which is propagated by using unsuitable features of data sets (e.g. p-values) or non-robust classifiers (e.g. LDA).
- The performance is drastically reduced with an increasing number of methods (in our case, the classification is

difficult already for 5 regression estimators).
- Metalearning attempts to perform the regression modelling automatically, without taking into account variable selection or interpretation of the model.
- Metalearning relies on statistical associations, but cannot capture causality.
- Metalearning can be justified merely as an assistive approach perhaps for a very specific set of data sets from one domain.

Acknowledgements

The project was supported by the project 17-01251S of the Czech Science Foundation.

References

[1] Brazdil, P., Giraud-Carrier, C., Soares, C., and Vilalta, E.: *Metalearning: Applications to data mining*. Springer, Berlin, 2009.

[2] Davies, L., and Gather, U.: Breakdown and groups. *Annals of Statistics* **33**, 3 (2005), 977–1035.

[3] Grabczewski, K.: *Meta-learning in decision tree induction.* Springer, Cham, 2014.

[4] Jurečková, J., Sen, P.K., and Picek, J.: *Methodology in robust and nonparametric statistics*. CRC Press, Boca Raton, 2012.

[5] Kalina, J.: On multivariate methods in robust econometrics, *Prague Economic Papers* **21** (2012), 69–82.

[6] Kalina, J.: Implicitly weighted methods in robust image analysis, *Journal of Mathematical Imaging and Vision* **44** (2012), 449–462.

[7] Kalina, J.: Least weighted squares in econometric applications, *Journal of the Applied Mathematics, Statistics and Informatics* **5** (2009), 115–125.

[8] Lichman, M. *UCI Machine Learning Repository*. Online [http://archive.ics.uci.edu/ml], 2013.

[9] Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibián-Barrera, M., Verbeke, T., Koller, M., Conceicao, E.L.T., and di Palma, M.A.: *robustbase: Basic Robust Statistics R package version 0.92-7,* 2016.

[10] Rokach, L., and Maimon, O.: *Data mining with decision trees.* World Scientific, New Jersey, 2008.

[11] Rousseeuw, P.J., and Leroy, A.M.: *Robust regression and outlier detection.* Wiley, New York, 1987.

[12] Smith-Miles, K., Baatar, D., Wreford, B., and Lewis, R.: Towards objective measures of algorithm performance across instance space, *Computers and Operations Research* **45** (2014), 12–24.

[13] Spaeth, H.: *Mathematical algorithms for linear regression.* Academic Press, Cambridge, 1991.

Optimal Value of Loans via Stochastic Programming

Vlasta Kaňková[1]

Abstract. A question of mortgage leads to serious and complicated problems of financial mathematics. On one side is a bank with an aim to have a "good" profit, on the other side is the client trying to invest money safely, with possible "small" risk. Let us suppose that a young married couple is in a position of client. Young people know that an expected and also unexpected unpleasant financial situation can happen. Many unpleasant financial situation can be caused by a random factor. Consequently stochastic methods are suitable to secure against them.
The aim of the suggested model is not only to state a maximal reasonable value of loans, but also to endure unpleasant financial period. To this end we employ stochastic optimization theory. A few suitable models will be introduced. The choice of the model depends on environment of the young people. Models will be with "deterministic" constraints, probability constraints, but also with stochastic dominance constraints. The suggested models will be analyzed both from the numerical point of view and from possible method solution based on data. Except static one–objective problem we suggest also multi–objective models.

Keywords: Loan, debtor, installments, stochastic programming, second order dominance constraints, probability constraints.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Let us construct and analyze a very simple financial model. To this end we consider a situation about a mortgage and instalments supposing a situation of young married couple. Young people wants to get own residence (a flat or a little house). Since they do not posses necessary financial resources, the bank sector offers them a mortgage. Of course bank can employ excellent experts to minimize their risk and maximize profit in dependence of debtor's position. The aim of our approach is to analyze the situation from the second side. In particular our aim is to investigate the possibilities of the debtors not only in dependence of their present–day situation, but also on their future private and subjective decision and on the possible "unpleasant" financial situation. In details, the aim is to suggest a method for a security of a "safe" loan and simultaneously to offer tactics to state a plausible environment for future time. Of course we suppose that our analysis is one of the first contribution to this situation. To this end we start with very standard situation of young people considered already in [6]. The young married couple decides to take loan of the value M. A question is how to choice value of M to be safe for them and simultaneously to secure them their wish of a comfortable flat.

To start with a responsible analyze of their situation we assume that a monthly income of young married people in a start point $t = 0$ is

$$Z_0 = U_0 + V_0, \quad \text{where} \quad U_0 \quad \text{is an income of husband and} \quad V_0 \quad \text{is an income of the wife.}$$

Evidently, this income can be divided into three parts Z_0^1, Z_0^2, Z_0^3, where Z_0^1 denotes means for a basic consumption, Z_0^2 denotes means that can be employed for a repayment of installments and Z_0^3 can be considered as an allocation to saving. Consequently

$$Z_0 = Z_0^1 + Z_0^2 + Z_0^3, \quad Z_0^1, Z_0^2 > 0, Z_0^3 \geq 0. \tag{1}$$

Supposing the annuity repayments, which is the most standard way of repaying the loan, we denote (as mentioned already above) by a symbol M the value of the loan, by m number of identical installments and by ζ the loan interest rate, then the identical installments $b(M) := b(M, \zeta)$ at time points $t = 1, 2, \ldots, m$ (see, e.g., [7] or [12]) are given by

$$b(M) := b(\zeta) = \begin{array}{ll} \frac{M\zeta}{1-v^m}, & \zeta \neq 0, \quad v = v(\zeta) = (1+\zeta)^{-1}, \\ \frac{1}{m}, & \zeta = 0. \end{array} \tag{2}$$

[1]Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, kankova@utia.cas.cz

It follows from the relations (1), (2) that (in the case when $\zeta \neq 0$) it is desirable (in "static" approach) the following inequality

$$\frac{M\zeta(1+\zeta)^m}{(1+\zeta)^m - 1} = Z_0^2 \tag{3}$$

to be fulfilled. Of course, this condition (in the extreme case) can be replaced by the inequality

$$\frac{M\zeta(1+\zeta)^m}{(1+\zeta)^m - 1} \leq Z_0^2 + Z_0^3. \tag{4}$$

If it is possible to assume that the relations (1), (2) will be fulfilled also in future, then the young people can take the loan equal to the maximal value M for which the inequality (3) (respective (4)) is fulfilled. However mostly it is necessary to assume that the financial situation of young married couple can change. For example: it is reasonable to assume that in some time period, say $(m_1,\, m_2)$, $0 < m_1 < m_2 < m$ the married couple plan to have a baby. According to this fact and to the social politics of the state the young people can assume the less income in this time. We construct a few mathematical models according to the client possibilities; furthermore we analyze them. To this end the theory of stochastic programming will be employed.

2 Simple Mathematical Models

2.1 Analysis of Situation

To construct mathematical models we suppose that the aim of the young people is to determine maximal "reasonable" safety loan M. To this end let

- $X_t \subset R^n$ ..... nonempty compact sets $t \in \{0,\, \ldots,\, m\}$,
- M ..... value of loan,
- m ..... number of identical installments,
- ζ ..... interest rate corresponding to the loan,
- Z_t ..... income of young married couple at time point $t \in \{0,\, 1,\, \ldots,\, m\}$,
- U_t ..... income of husband at time point $t \in \{0,\, 1,\, \ldots,\, m\}$,
- V_t ..... income of wife at time point $t \in \{0,\, 1,\, \ldots,\, m\}$,
- Z_t^1 ..... means determined for basic consumption at time point $t \in \{0,\, 1,\, \ldots,\, m\}$,
- Z_t^2 ..... means determined for repayment of installment at time point $t \in \{0,\, 1,\, \ldots,\, m\}$,
- Z_t^3 ..... allocation (maybe random) for saving in time point $t \in \{0,\, 1,\, \ldots,\, m\}$,
- $\langle m_1,\, m_2\rangle$ ..... time interval in which income of wife is supposed to be smaller,
- $\xi_{t,\,j}$, ... random returns at time t and assets j, $t \in \{0,\, 1, \ldots,\, m\}$, $j \in \{1,\, 2\}$,
- $J = \{1,\, 2\}$ system of assets,
- $x_{t,\,j}$ ..... decision variables, $t \in \{0,\, 1, \ldots,\, m\}$, $j \in \{1,\, 2\}$,
- $x_t = [x_{t,\,1},\, x_{t,\,2}]$, $t = 0,\, 1,\, \ldots,\, m$,
- $\xi_t = [\xi_{t,\,1},\, \xi_{t,\,2}]$, $t = 0,\, 1,\, \ldots,\, m$,
- $g_t = \xi_{t,\,1}x_{t,\,1} + \xi_{t,\,2}x_{t,\,2}$, $t = 0,\, \ldots,\, m$,
- $Y_t = \frac{Z_t^3}{2}\xi_{t,\,1} + \frac{Z_t^3}{2}\xi_{t,\,2}$, $t = 0,\, \ldots,\, m$,
- F ..... distribution function covering all random values that occur in the corresponding model,
- $\mathcal{Z}$ ..... support corresponding to F.

Remark. We suppose (without loss of generality) that young people can the amount Z_t^3 (at every time point $t \in \langle 0,\, m\rangle$) invest only in two assets $J = \{1,\, 2\}$. Of course, in a real–life situation mostly more possibilities exist.

To analyze the situation we suppose that M fulfils the relation

$$\frac{M\zeta(1+\zeta)^m}{(1+\zeta)^m - 1} \leq Z_t^2 \quad \text{for} \quad t = 0,\, \ldots m_1 - 1,\, m_2 + 1,\, \ldots,\, m.$$

Consequently if

$$\frac{M\zeta(1+\zeta)^m}{(1+\zeta)^m - 1} \leq Z_t^2 \quad \text{also for every} \quad t \in \langle m_1,\, m_2\rangle, \tag{5}$$

then everything is OK. (This can happen, if for example one of the following situation happen: M is small in comparison with income of young people; the husband have two positions in the time interval $\langle m_1,\, m_2\rangle$; the wife can work home; parents helps). But usually the situation is not such pleasant and the following inequality

$$\frac{M\zeta(1+\zeta)^m}{(1+\zeta)^m - 1} > Z_t^2 \quad \text{can happen for some} \quad t \in \langle m_1,\, m_2\rangle. \tag{6}$$

Evidently, the trouble starts in the case (6). Young people can protect against them. Especially they can save at the time points $t = 0,\ 1,\ \ldots,\ m_1 - 1$ means Z_t^3 to be fulfilled the inequality

$$\frac{(m_2 - m_1 + 1)M[\zeta(1+\zeta)^m]}{(1+\zeta)^m - 1} \leq (m_2 - m_1 + 1)[Z_0^2 - Z_{m_1}^2] + \sum_{t=0}^{m_1-1} Z_t^3, \tag{7}$$

(under the assumption $Z_t^2 = Z_0^2,\ t = 0,\ \ldots,\ m_1 - 1;\ Z_t^2 = Z_{m_1}^2,\ t = m_1,\ \ldots,\ m_2$. If the last inequality (7) is fulfilled, then they endure the time period $\langle m_1,\ m_2\rangle$ without financial troubles.

2.2 Model Construction

Till now we have assumed that Z_t^3 is deterministic value. It is known that a random component very often exists in salary. Just this component is suitable for saving. We consider in our analysis both cases: deterministic and random. However for simplicity (without loss of generality) we consider only special case $m_1 = 2, m_2 = 4$; m given by the relation (2). First we consider a completely deterministic model:

I. The young people can invest the deterministic amounts $Z_0^3,\ Z_1^3$ into two deterministic assets to obtain

$$\begin{aligned}
&\text{in the first year the value} && c_{0,1}x_{0,1} + c_{0,2}x_{0,2}\\
&\text{under the assumptions} && x_{0,1} + x_{0,2} \leq Z_0^3, \quad x_{0,1}, x_{0,2} \geq 0, \quad x_0 \in X_0,\\
&\text{in the second year the value} && c_{1,1}x_{1,1} + c_{1,2}x_{1,2}\\
&\text{under the assumptions} && x_{1,1} + x_{1,2} \leq Z_1^3, \quad x_{1,1}, x_{1,2} \geq 0, \quad x_1 \in X_1,
\end{aligned}$$

where $c_{i,\,j},\ i = 0,\ 1,\ j = 1,\ 2$ are deterministic constants.
We assume in this model that the profit obtained at time $t = 0$ can not influence the invested amount at time $t = 1$.
Evidently, it is desirable (for young people) fulfilling of the relation

$$\frac{(m_2 - m_1 + 1)M[\zeta(1+\zeta)^m]}{(1+\zeta)^m - 1} \leq (m_2 - m_1 + 1)[Z_0^2 - Z_{m_1}^2] + \sum_{i=0}^{1}[c_{i,1}x_{i,1} + c_{i,2}x_{i,2}]. \tag{8}$$

Consequently, supposing $m_1 = 2,\ m_2 = 4;\ Z_0^2 = Z_1^2;\ Z_{m_1}^2 = Z_t^2, t \in \langle m_1,\ m_2\rangle;\ c_{i,\,j},\ i,\ j = 1,\ 2$ deterministic, we obtain a deterministic optimization model:

$$\text{Find} \qquad \max M \tag{9}$$

under the constraints

$$\begin{aligned}
x_{0,1} + x_{0,2} \quad &\leq Z_0^3, \quad x_{0,1}, x_{0,2} \geq 0, \quad x_0 \in X_0.\\
x_{1,1} + x_{1,2} \quad &\leq Z_1^3, \quad x_{1,1}, x_{1,2} \geq 0, \quad x_1 \in X_1,\\
\tfrac{M\zeta(1+\zeta)^m}{(1+\zeta)^m - 1} \quad &\leq Z_t^2 \quad \text{for} \quad t = 0, \ldots m_1 - 1,\ m_2 + 1,\ \ldots,\ m,\\
\tfrac{3M[\zeta(1+\zeta)^m]}{(1+\zeta)^m - 1} \quad &\leq 3[Z_0^2 - Z_2^2] + \sum_{i=0}^{1}[c_{i,1}x_{i,1} + c_{i,2}x_{i,2}].
\end{aligned}$$

The problem (9) is a problem of linear programming. Consequently, it can be analyzed and solved employing the theory of linear programming.

II. We consider again $Z_t^3,\ t = 0,\ 1,\ \ldots,\ m$ deterministic. However, in the difference to the case I, young people can invest the value Z_t^3 into two assets with random returns $\xi_{t,1},\ \xi_{t,2},\ t = 0,\ 1$. Consequently it is necessary to determine $x_{0,1},\ x_{0,2},\ x_{1,1},\ x_{1,2}$ fulfilling the relations

$$\begin{aligned}
&&& x_{0,1} + x_{0,2} \leq Z_0^3, \quad x_{0,1},\ x_{0,2} \geq 0,\\
&&& x_{1,1} + x_{1,2} \leq Z_1^3, \quad x_{1,1},\ x_{1,2} \geq 0\\
&\text{to obtain random values} && g_0 := g_0(x_0,\ \xi_0) = \xi_{0,1}x_{0,1} + \xi_{0,2}x_{0,2},\\
&&& g_1 := g_1(x_1,\ \xi_1) = \xi_{1,1}x_{1,1} + \xi_{1,2}x_{1,2}.
\end{aligned}$$

Evidently, it is possible also to define random values $Y_0,\ Y_1$ by

$$\begin{aligned} Y_0 : Y_0(\xi_0) &= \tfrac{Z_0^3}{2}\xi_{0,1} + \tfrac{Z_0^3}{2}\xi_{0,2}, \\ Y_1 := Y_1(\xi_1) &= \tfrac{Z_1^3}{2}\xi_{1,1} + \tfrac{Z_1^3}{2}\xi_{1,2}. \end{aligned} \tag{10}$$

Employing the theory of the stochastic dominance [11] it is "reasonable" to determine $x_{0,1},\ x_{0,2},\ x_{1,1},\ x_{1,2}$ such that

$$\begin{aligned} & F_{g_0} \succeq_1 F_{Y_0}, \qquad F_{g_1} \succeq_1 F_{Y_1}, \\ \text{or} \quad & F_{g_0} \succeq_2 F_{Y_0}, \qquad F_{g_1} \succeq_2 F_{Y_1}. \end{aligned} \tag{11}$$

The first relation in (11) is known as a stochastic dominance of the first order; the second relation is known as stochastic dominance of the second order. To define stochastic dominance of the second order it is necessary to assume that the first moments of the random values $g_0(x_0,\ \xi_0),\ g_1(x_1,\ \xi_1),\ Y_0,\ Y_1,$ exist for $x_0 \in X_0,\ x_1 \in X_1$. (More about the definition of the stochastic dominance can be find, e.g., in [11].)

Of course, it is also desirable (for young people) in this case the fulfilling of the relation

$$\frac{3M[\zeta(1+\zeta)^m]}{(1+\zeta)^m - 1} \le 3[Z_0^2 - Z_2^2] + \sum_{i=0}^{1}[\xi_{i,1}x_{i,1} + \xi_{i,2}x_{i,2}]. \tag{12}$$

However, the inequality (12) depends on the random elements $\xi_{i,\,j},\ i = 0,\ 1,\ j = 1,\ 2$. Consequently, the "sense" of this inequality has to be defined. We consider it in probability.

Consequently, we can obtain the following optimization model depending on a probability measure:

$$\text{Find} \qquad \max M \tag{13}$$

under the constraints

$$\begin{aligned} x_{0,1} + x_{0,2} &\le Z_0^3, \quad x_{0,1},\ x_{0,2} \ge 0, \\ x_{1,1} + x_{1,2} &\le Z_1^3, \quad x_{1,1}, x_{1,2} \ge 0, \\ \tfrac{M\zeta(1+\zeta)^m}{(1+\zeta)^m - 1} &\le Z_t^2 \quad \text{for} \quad t = 0,\ 1,\ 5,\ \ldots,\ m, \end{aligned} \tag{14}$$

$$P_F\{\frac{3M[\zeta(1+\zeta)^m]}{(1+\zeta)^m - 1} \le 3[Z_0^2 - Z_{m_1}^2] + \sum_{i=0}^{1}[\xi_{i,1}x_{i,1} + \xi_{i,2}x_{i,2}]\} \ge 1 - \varepsilon, \quad \varepsilon \in (0,\ 1), \tag{15}$$

$$\begin{aligned} \mathsf{E}_F(u - g_0(x_0,\ \xi_0))^+ &\le (u - Y_0(\xi_0))^+, \quad u \in R^1,\ x_0 \in X_0, \\ \mathsf{E}_F(u - g_1(x_1,\ \xi_(1)))^+ &\le (u - Y_1(\xi_1))^+, \quad u \in R^1,\ x_1 \in X_1. \end{aligned} \tag{16}$$

The equivalence of the relation (11) and (16) has been proven by Ruszczynski, see e.g., [11].

The constraints (14) are linear, constraints (16) under general conditions are convex. However (from the numerical point of view) the constraint (15) can be a problem.

III. Deterministic $Z_t^3,\ t = 0,\ 1,\ \ldots,\ m$ can be replaced by random values with probability one to be non negative.

We assume that young people can random amounts $Z_0^3,\ Z_1^3$ invest into two assets to obtain:

in the original year the value	$\xi_{0,1}x_{0,1} + \xi_{0,2}x_{0,2}$
under the assumptions	$x_{0,1} + x_{0,2} \le Z_0^3, \quad x_{0,1}, x_{0,2} \ge 0, \quad x_0 \in X_0,$

in the second year the value	$\xi_{1,1}x_{1,1} + \xi_{1,2}x_{1,2}$
under the assumptions	$x_{1,1} + x_{1,2} \le Z_1^3, \quad x_{1,1}, x_{1,2} \ge 0, \quad x_1 \in X_1.$

We assume in this case that the profit obtained in the time $t = 0$ can not influence the invested amount at the time $t = 1$.

Evidently, it is desirable in this case the fulfilling of the relation

$$\frac{3M[\zeta(1+\zeta)^m]}{(1+\zeta)^m-1} \leq 3[Z_0^2 - Z_2^2] + \sum_{t=0}^{1}[\xi_{t,1}x_{t,1} + \xi_{t,2}x_{t,2}] \tag{17}$$

and, simultaneously, the constraints with random factors

$$x_{0,1} + x_{0,2} \leq Z_0^3, \quad x_{1,1} + x_{1,2} \leq Z_1^3.$$

We consider all these constraints with the random factors in probability. Consequently we obtain stochastic optimization problem:

$$\text{Find} \qquad \max M \tag{18}$$

under the system of constraints

$$\frac{M\zeta(1+\zeta)^m}{(1+\zeta)^m-1} \leq Z_0^2, \quad t = 0, 1, \ 5, ...m, \tag{19}$$

$$P_F\{x_{t,1} + x_{t,2} \leq Z_t^3\} \quad \geq \quad 1 - \varepsilon_t, \quad \varepsilon_t \in (0,1), \ x_{t,1}, \ x_{t,2} \geq 0, \quad t = 0, \ 1, \ 5, ...m, \tag{20}$$

$$\begin{aligned} P_F\{\tfrac{3M[\zeta(1+\zeta)^m}{(1+\zeta)^m-1}\} \quad &\leq \quad 3[Z_0^2 - Z_2 + \sum_{i=0}^{1}[\xi_{t,1}x_{t,1} + \xi_{t,2}x_{t,\,2}]\} \geq 1 - \varepsilon_0, \\ &\varepsilon_0 \in (0, \ 1). \end{aligned} \tag{21}$$

In this case the constraints (19) is linear, the constraint (20) can be rewritten into linear conditions [4]. Difficulty arises only with the constraint (20); the same as in the problem II.

Remark. In all introduced cases we consider only the problem of choice the value M. However, surely it is very reasonable and suitable to maximize profit obtained by an investment at all time point in the interval $\langle 0, \ m\rangle$ or in the time interval $\langle m_2 + 1, \ m\rangle$. Because the profit is a random value, it is reasonable to look for optimal solution with respect of the mathematical expectation. Evidently in this case we obtain two–objective optimization problem.

3 Conclusion

In the last decades many people try to gain their own residence. Since they do not posses sufficient means, the bank sector offer them the loan. The aim of this contribution is to give a preliminary analysis of their situations and possible responsible behaviour. In the paper a very simple stochastic optimization problems have been introduced. However the model has been constructed to guarantee only very small "risk" for young people. However on the other side only completely deterministic model can be solved by classical numerical methods. The other models obtained do not contain convex constraints.

But replacing theoretical distribution by empirical one we obtain "good" estimates of the original problem. To more details see the literature about empirical estimates e.g.,[1],[2], [4], [5], [8], [9], [10].

Acknowledgements

This work was supported by the Czech Science Foundation under grant 15-10331S.

References

[1] Dai, L, Chen, C. H. and Birge,J. R.: Convergence properties of two–stage stochastic programming. *J. Optim. Theory Appl.* **106** (2000), 489–509.

[2] Dvoretzky, A., Kiefer, J. and Wolfowitz, J.: Asymptotic Minimax character of the Sample Distribution Function and the Classical Multinomial Estimate. *Ann. Math. Statist.* **56** (1956), 642–669.

[3] Frank, H. and Glass, A.J.: *Microeconomics and Behavior.* Mc Graw–Hill, New York 1991.

[4] Kaňková, V.: On the stability in stochastic programming: the case of individual probability constraints. *Kybernetika* **33** (1997), 5, 525–546.

[5] Kaňková, V. and Houda, M.: Thin and Heavy Tails in Stochastic Programming. *Kybernetika* **51** (2015), 3, 433–455.

[6] Kaňková, V.: A Note of Optimal Value of Loans. In: *Proceedings of the 34th International Conference Mathematical Methods of Economics 2016* (A. Kocourek and M. Vavroušek, eds.), Technical Univesity Liberec, Liberec 2016, 371–376.

[7] Luderer, B., Nollau, V. and Vetters, K.: *Mathematical Formulas for Economists.* Springer Science & Media, third edition, 2006.

[8] Pflug, G. Ch.: Stochastic Optimization and Statistical Inference. In: *Stochastic Programming, Handbooks in Operations Research and Management Science, Vol. 10* (A. Ruszczynski and A. A. Shapiro, eds.), Elsevier, Amsterdam 2003, pp. 427–480.

[9] Römisch, W.: Stability of Stochastic Programming Problems. In: *Stochastic Programming, Handbooks in Operations Research and Management Science, Vol 10* (A. Ruszczynski and A.A. Shapiro, eds.), Elsevier, Amsterdam 2003, pp. 483–554.

[10] Shapiro, A.: Quantitative stability in stochastic programming. *Math. Program.* **67** (1994), 99–108.

[11] Shapiro, A., Dentcheva, D. and Ruszczynski, A.: *Lectures on Stochastic Programming (Modeling and Theory).* Published by Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia 2009.

[12] Šmíd, M. and Dufek, J.: Multi–period Factor Model of Loan Portfolio (July 10, 2016). Available at SSRN:http://ssrn.com/abstract=2703884 or http://dx.doi.org/10.2139/ssrn.2703884.

EWMA Based AOQL Variables Sampling Plans and Cost Models

Nikola Kaspříková[1]

Abstract. Rectifying acceptance sampling plans minimizing the mean inspection cost per lot of the process average quality were designed by Dodge and Romig for the inspection by attributes. Plans for the inspection by variables were then proposed and such plans may be more economical than the corresponding attributes sampling plans. The recently proposed plans using EWMA statistic may lead to further improvements in the inspection cost. We recall the design of the EWMA-based rectifying AOQL sampling plans minimizing the mean inspection cost per lot of process average quality and perform the evaluation of the comparative economic efficiency, using a simple economic model which considers differences in both variable and fixed costs. The fixed cost difference is considered in the evaluation since the decision to implement the plan for the inspection by variables and attributes may bring higher fixed cost in some cases in practice in comparison with performing just the attribute inspection due to presumably more difficult preparation of the inspection procedure. The R software extension package is used for the calculation of the plans.

Keywords: acceptance sampling, cost optimization, AOQL.

JEL classification: C44
AMS classification: 62D99

1 Introduction

Sampling inspection is a basic tool of industry statistics. The acceptance sampling plans are designed and used to improve the economic efficiency of the business process. There exist many types of such acceptance sampling plans. New rectifying average outgoing quality limit (AOQL) plans for sampling inspection by variables based on the exponentially weighted moving average (EWMA) statistic have been recently introduced. [6] addresses the case of known standard deviation. The EWMA-based Lot Tolerance Percent Defective (LTPD) plans for the unknown standard deviation case are introduced in [7]. This paper addresses the design of EWMA-based AOQL plans for the unknown standard deviation case and focuses on new approach to the comparative economic efficiency evaluation of the plans. The R software extension package [5] which implements the algorithms for the calculation of the plans and which has been published on the Comprehensive R Archive Network can be used for the calculation of the plans.

The AOQL sampling plans minimizing the mean inspection cost per lot of process average quality when the remainder of rejected lots is inspected were originally designed by Dodge and Romig (see e.g. [2]) for the inspection by attributes. Plans for the inspection by variables and for the inspection by variables and attributes (all items from the sample are inspected by variables, the remainder of rejected lots is inspected by attributes) were then proposed and it was shown that these plans are in many situations more economical than the corresponding Dodge-Romig attribute sampling plans. The AOQL plans for inspection by variables and attributes have been introduced in [9], using approximate calculation of the plans. Exact operating characteristic, using non-central t distribution, has been later implemented for the calculation of the plans in the LTPDvar package [5]. The operating characteristics used for these plans are discussed by Jennett and Welch in [3] and by Johnson and Welch in [4]. It has been shown that these plans are in many situations superior to the original attribute sampling plans and similar results have been obtained for the LTPD plans, the analysis is provided in [8]. The recent development of acceptance sampling plans includes the work by Aslam et al. in [1] where the EWMA statistic is used for a design of the (p_1, p_2) sampling plans, i.e. sampling plans which satisfy the requirement to control the producer's risk and the consumer's risk. Using the EWMA statistic enables some savings in the cost of inspection as it allows using information on the quality in the previous lots. With the aim of obtaining further savings in the cost of inspection, the new AOQL plans for the inspection by variables and attributes, designed to use the EWMA statistic, have been proposed in [6].

The structure of this paper is as follows: first, the design of the AOQL sampling plans for the inspection by attributes, as introduced by Dodge and Romig (see [2]), is recalled. Then we introduce the AOQL variables sampling plans based on the usage of the EWMA statistic in the inspection procedure. We design a single sampling

[1]University of Economics in Prague, Department of Mathematics,
Nám. W. Churchilla 4, Praha, Czech Republic, nb33@tuliany.cz

plan. The EWMA statistic based AOQL plans have been so far discussed for the case of the known standard deviation - see [6], but in many situations in the business practice, the unknown standard deviation case might be more appropriate, so we introduce such plans in the introductory section of the paper. Then the newly designed plans are calculated, using the LTPDvar package (see [5]) and a comparative economic efficiency evaluation is discussed using a simple cost model which compares the economic efficiency of the new plans with the attribute sampling plans.

1.1 Attributes inspection plans

For the case that each inspected item is classified as either good or defective (the acceptance sampling by attributes), Dodge and Romig (see [2]) consider sampling plans $(n,\ c)$ which minimize the mean number of items inspected per lot of process average quality, assuming that the remainder of the rejected lots is inspected

$$I_s = N - (N - n)\cdot L(\bar{p}; n, c) \tag{1}$$

under the condition

$$\max_{0<p<1} AOQ(p) = p_L. \tag{2}$$

The notation in equations (1) and (2) is as follows:
N is the number of items in the lot (the given parameter),
$\bar{p}$ is the process average fraction defective (the given parameter),
p_L is the average outgoing quality limit (the given parameter, denoted AOQL),
n is the number of items in the sample ($n < N$),
c is the acceptance number (the lot is rejected when the number of defective items in the sample is greater than c),
$L(p)$ is the operating characteristic (the probability of accepting a submitted lot with the fraction defective p).

The function AOQ is the *average outgoing quality*, $AOQ(p)$ is the mean fraction defective after inspection when the fraction defective before inspection was p. The AOQ function is continuous in [0,1] and reaches minimum value 0 for $p = 0$ and for $p = 1$. For a more detailed discussion of the properties of the AOQ function, see [2]. The average outgoing quality (where all defective items found are replaced by good ones) is approximately

$$AOQ(p) = \left(1 - \frac{n}{N}\right)\cdot p \cdot L(p; n, c). \tag{3}$$

Therefore the condition (2) can be rewritten as

$$\max_{0<p<1} \left(1 - \frac{n}{N}\right)\cdot p \cdot L(p; n, c) = p_L. \tag{4}$$

The condition (2) protects the consumer against having an average outgoing quality higher than p_L (the chosen value), regardless of what the fraction defective p is before inspection.

1.2 Sampling plans for the inspection by variables

We introduce the AOQL plans for the inspection by variables and attributes based on the EWMA statistic for the case of the unknown standard deviation. The design is similar to the design of the recently introduced plans for the known standard deviation case - see [6]. But the operating characteristic used in [7] for the design of the LTPD plans is applied here. The new AOQL plans are designed under the following assumptions: The measurements of a single quality characteristic X are independent and identically distributed normal random variables with parameters μ and σ^2. We consider the unknown σ case. For the quality characteristic X, either an upper specification limit U (the item is defective if its measurement exceeds U), or a lower specification limit L (the item is defective if its measurement is smaller than L), is given.

For the design of the rectifying AOQL plans minimizing the mean inspection cost per lot of the process average quality we shall use a procedure based on the EWMA statistic. The procedure is as follows: draw a random sample of n items from the lot and compute the sample mean $\bar{x}$, sample standard deviation s and the statistic Z at time t as

$$Z_t = \lambda\bar{x} + (1 - \lambda)Z_{t-1}, \tag{5}$$

where λ is a smoothing constant (between 0 and 1).

Accept the lot if

$$\frac{U - Z_t}{\sigma} \geq k \quad \text{or} \quad \frac{Z_t - L}{\sigma} \geq k. \tag{6}$$

The operating characteristic used here is (see e.g. [1])

$$L(p) = \Phi(u_{1-p}c_4 - k)\sqrt{\frac{1}{\frac{\lambda}{n(2-\lambda)} + k^2(1 - {c_4}^2)}}, \tag{7}$$

where

$$c_4 = \sqrt{(2/(n-1))}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}. \tag{8}$$

The plan parameters (n, k) are to be determined so that the plan has optimal economic characteristics and satisfies the requirement (2), when (7) is used as the operating characteristic.

Regarding the economic optimality, Klufa in [9] uses the economic model, used for a more detailed evaluation of the sampling plans in [8] and searches for the acceptance plan (n, k), minimizing the mean inspection cost per lot of the process average quality C_{ms} under the condition (2). We are going to use this model for the design of the plans too. The inspection cost per lot, assuming that the remainder of the rejected lots is inspected by attributes (the inspection by variables and attributes), is $n\, c_m^*$, with the probability $L(p; n, k)$, and $[n\, c_m^* + (N-n)\, c_s^*]$ with the probability $[1-L(p; n, k)]$, where c_s^* is the cost of the inspection of one item by attributes, and c_m^* is the cost of the inspection of one item by variables. The mean inspection cost per lot of the process average quality is then

$$C_{ms} = n{\cdot}c_m^* + (N-n){\cdot}c_s^*{\cdot}[1 - L(\bar{p}; n, k)]. \tag{9}$$

Let us denote

$$c_m = \frac{c_m^*}{c_s^*}. \tag{10}$$

Instead of C_{ms} we will look for the acceptance plan (n, k) minimizing

$$I_{ms} = n{\cdot}c_m + (N-n){\cdot}[1 - L(\bar{p}; n, k)] \tag{11}$$

(both functions C_{ms} and I_{ms} have a minimum for the same acceptance plan ($C_{ms} = I_{ms}{\cdot}\, c_s^*$)) under the condition (2).

For the discussion of the c_m parameter, see [8]. The value of this parameter must be estimated based on the real cost calculation in practice. Usually it is $c_m > 1$.

2 Calculation and evaluation of the plans with LTPDvar package

2.1 Calculation of the plan

The AOQL acceptance sampling plan based on the EWMA model for sampling inspection by variables when the remainder of rejected lots is inspected by attributes, as implemented in [5] will be calculated in the example below. The solution for the unknown σ case will be searched for, using the operating characteristic given by (7). The resulting sampling plan will then be evaluated with regard to the economic characteristics and compared with the corresponding AOQL attributes sampling plan as discussed in [2].

Example 1. A lot with $N = 3500$ items is considered in the acceptance procedure. The average outgoing quality limit is given to be p_L = 0.015. It is known that the average process quality is $\bar{p}$ = 0.01. A cost of inspecting an item by variables is two times higher than the cost of inspecting the item by attributes, so the parameter c_m equals 2. Find the AOQL acceptance sampling plan for sampling inspection by variables when remainder of rejected lots is inspected by attributes, using the EWMA statistic with smoothing constant $\lambda = 0.9$.

The plan can be calculated using the functions available in the LTPDvar package for the R software [10], see the documentation of the package for a more detailed description. The implementation is based on the following principles: the k corresponding to particular sample size n is obtained using the condition (2) and the cost-optimizing n is searched for using numerical methods, where the approximate solution, which is calculated first, is used for setting the bounds of the intervals to be searched in.

The resulting plan may be obtained using the following call in the LTPDvar package:

```
planAOQL(N=3500, pbar=0.01, pL=0.015, cm=2,method="ewma2", lam=0.9)
```

and the output of the calculation is as follows.

```
An object of class "ACSPlan"
Slot "n":
[1] 65

Slot "k":
[1] 1.91832
```

So the resulting solution is $n = 65$, $k = 1.91832$.

For the values of the input parameters given in our problem, there is plan (165, 4) for the acceptance sampling by attributes in [2].

2.2 Comparative economic evaluation of the plan

For the comparison of these two plans from Example 1 - the EWMA plan for inspection by variables and attributes ($n = 65$, $k = 1.91832$) and the Dodge-Romig plan ($n = 165$, $c = 4$) from an economic point of view we can use a simple economic model of cost difference (measured in the c_s^* units) in which we consider the difference of the mean inspection costs per lot of the process average quality of the particular plans and the effect of the difference in fixed costs:

$$c = I_{ms}(65, 1.91832) - I_s(165, 4) + f, \tag{12}$$

where f is the increase in fixed cost (measured in c_s^*) in case that we opt for the plan for the inspection by variables and attributes. The decision to implement the plan for the inspection by variables and attributes may bring higher fixed cost in some cases in practice in comparison with performing just the attribute inspection due to presumably more difficult preparation of the inspection procedure, including a considerations regarding whether or not the assumptions for the application of the variables sampling plan are met and so on.

The I_{ms} function is implemented in the LTPDvar package too and (for plan $n = 65$, $k = 1.91832$) may be called as shown below.

```
Ims(n= 65, k=1.91832, N= 3500, pbar=0.01, cm=2, type="ewma2", lam=0.9)
```

In this case we obtain $I_{ms} = 199.9$. The cost measure c then represents the cost difference (measured in the c_s^* units) in the mean inspection cost per lot of the process average quality in case that the EWMA based plan for the inspection by variables and attributes is used in place of the corresponding attribute sampling plan. The guidelines for the application and interpretation of the measure are as follows: in case that the resulting c value is negative, then one should preferably choose the plan for the inspection by variables and attributes.

Let's suppose now that in our example situation it is $f = 30\, c_s^*$, and let's have a look at the comparative economic efficiency of the variable sampling plan in such a case. To resolve whether or not one should opt for the sampling plan for the inspection by variables, the value of the cost measure c could be calculated.

Then for the input values in our example situation we get $c = -10.4$, which shows (since the resulting value for c is below 0) that there can be expected some savings in the mean inspection cost per lot of the process average quality if the variables sampling plan is used in place of the corresponding attribute inspection plan.

Nevertheless, it also could be the case that the value of f is higher in some situations in the business practice. Let's perform the evaluation using the cost measure c for such case.

For example if the value of f is as high as $70\, c_s^*$, then we get the resulting value of the cost measure c as high as 30.4. And in such case, one should rather not opt for the variables sampling plan and use the attribute sampling plan as introduced by Dodge and Romig instead, since we obtained the resulting value of c greater than zero.

It may be sometimes rather difficult to get the precise values for all the input values of the particular cost characteristics (like c_m, c_s, f) needed to be able to perform the economic evaluation of the plans which are considered. But this is unfortunately quite common situation in many other fields in business where statistical methods are applied, like the classification problems solved by modern data mining techniques (where one may not be always sure regarding the cost of particular misclassification error), to mention just one. But it may pay to perform at least some attempt to evaluate the alternatives considered, using some reasonably precise estimates of the real cost characteristics values.

3 Conclusion

The EWMA-based unknown standard deviation case AOQL plans for the sampling inspection by variables when the remainder of the rejected lots is inspected by attributes minimizing the mean inspection cost per lot of the

process average quality may in many cases in the business practice bring significant economic advantages over the corresponding attribute inspection plans. The improvement in economic efficiency can be assessed using a simple economic model considering the difference in variable and fixed cost between the variables sampling plan and the attributes sampling plan. The LTPDvar package for the R computing environment can be conveniently used for the calculation of the optimal acceptance sampling plan.

Acknowledgements

This paper has been produced with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Aslam, M., Azam, M., and Jun, C.: A new lot inspection procedure based on exponentially weighted moving average, *International Journal of Systems Science* **46** (2015), 1392–1400.

[2] Dodge, H., and F., Romig, H. G.: *Sampling Inspection Tables: Single and Double Sampling*. John Wiley, New York, 1998.

[3] Jennett, W. J., and Welch, B. L.: The Control of Proportion Defective as Judged by a Single Quality Characteristic Varying on a Continuous Scale, *Supplement to the Journal of the Royal Statistical Society* **6** (1939), 80–88.

[4] Johnson, N. L., and Welch, B. L..: Applications of the Non-central t distribution, *Biometrika* **38** (1940), 362–389.

[5] Kasprikova, N.: *LTPDvar: LTPD and AOQL plans for acceptance sampling inspection by variables.* R package version 1.2. http://CRAN.R-project.org/package=LTPDvar, 2016.

[6] Kasprikova, N.: Calculation and evaluation of new AOQL single sampling plans for inspection by variables. In: *Mathematical Methods in Economics 2015*, University of West Bohemia, Cheb, 2015, 349–353.

[7] Kasprikova, N.: Rectifying LTPD Plans Minimizing Mean Inspection Cost Per Lot of Process Average Quality Using EWMA Statistics. *Journal of Testing and Evaluation* **4** (2017), DOI: 10.1520/JTE20150259.

[8] Kasprikova, N., and Klufa, J.: AOQL sampling plans for inspection by variables and attributes versus the plans for inspection by attributes, *Quality technology and quantitative management* **6** (2015), 133–142.

[9] Klufa, J.: Dodge-Romig AOQL single sampling plans for inspection by variables, *Statistical Papers* **38** (1997), 111–119.

[10] R Core Team: *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2017. URL http://www.R-project.org

Modified Transportation Problem

Renata Kawa[1], Rafał Kucharski[2]

Abstract. The paper considers a modified version of a transportation problem. Classical formulation aims at finding a transportation plan with the minimal cost, given supplies and demands of sources and destinations and a matrix of unit transportation costs. However, in our modification some roads between sources and destinations are not allowed. In practice, one may encounter this kind of problem in case of embargo, war or natural disaster. As a result, classical algorithms such as North-West Corner Method are no longer directly applicable, and there is no guarantee for the existence of a feasible solutions.
The main result of the paper is a necessary and sufficient condition for the existence of feasible solutions of this modified transportation problem. Proof of sufficiency provides an algorithm for finding a feasible solution, based on „divide and conquer" strategy. Paper also includes a simple example and some efficiency considerations. The result has already found an application in cryptography.

Keywords: transportation problem, forbidden roads, feasible solution

JEL classification: C44 , R42
AMS classification: 90C08 , 90C27 , 90C35

1 Introduction

In this paper we consider a modified version of a transportation problem: we aim at finding a transportation plan with the minimal cost, given supplies and demands of sources and destinations and a matrix of unit transportation costs. In a classical formulation, all shipment routes, between all sources and destination are possible. The modification we consider here is that some routes may not be allowed. In practice, one may encounter this kind of bounds in case of embargo, war or natural disaster. As a result, classical algorithms such as North-West Corner Method are no longer directly applicable, and there is no guarantee for the existence of a feasible solutions.

The main result of the paper is a necessary and sufficient condition for the existence of feasible solutions of this modified transportation problem. Proof of sufficiency provides an algorithm for finding a feasible solution, resembling a „divide and conquer" strategy. The paper also includes a simple example and some efficiency considerations. We also provide an example of application of this result in cryptography.

2 Statement of a problem

2.1 Classical Transportation Problem

Let us start with reminding the classical transportation problem (CTP). Suppose that we have a finite set of m sources (e.g. mines) producing products (e.g. iron ore), and a finite set of n destinations (e.g. factories which consume the iron ore). Moreover we are given the unit costs of shipment of the products from sources to destinations. The purpose is to find a scheme of product transport from sources to destinations with the minimal costs, assuming the shipment cost is linear (proportional to the quantity of product transported from a given source to a given destination). We are also bound by a supply of each source, and a demand of each destination. Let us adopt the following notations:

- m – a number of sources,
- n – a number of destinations,
- $\mathbf{S} = \{S_1, \ldots, S_m\}$ – a set of sources,
- $\mathbf{D} = \{D_1, \ldots, D_n\}$ – a set of destinations,
- s_i – a supply of the i-th source, $i = 1, \ldots, m$,
- d_j – a demand of the j-th destination, $j = 1, \ldots, n$,
- c_{ij} – transport costs of one unit of the product from the i-th source to the j-th destination,
- x_{ij} – a number of units of the product transported from the i-th source to the j-th destination.

[1]University of Silesia in Katowice, Institute of Mathematics, Bankowa 12, 40-007 Katowice, Poland, renata.kawa@us.edu.pl
[2]University of Economics in Katowice, Faculty of Management, 1 Maja 50, 40-287 Katowice, Poland, rafal.kucharski@ue.katowice.pl

With the above notation CTP may be formulated in the following way: find $x_{ij} \in \mathbb{R}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, such that

$$\begin{cases} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \to \min, \\ \sum_{j=1}^{n} x_{ij} = s_i \text{ for } i = 1, \ldots, m, \\ \sum_{i=1}^{m} x_{ij} = d_j \text{ for } j = 1, \ldots, n, \\ x_{ij} \geq 0, \text{ for } i = 1, \ldots, m, \ j = 1, \ldots, n. \end{cases} \tag{CTP}$$

We say that the CTP problem is balanced, if total demand equals total supply:

$$\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j,$$

If the above condition is not met, we can balance the problem introducing an artificial source or an artificial destination.

For a balanced CTP a feasible solution always exists, what is an obvious consequence of the following algorithm.

1. Start with $i := 1$, $j := 1$.
2. If $s_i > d_j$, then
   - $x_{ij} := d_j$, $x_{kj} = 0$ for $k > i$,
   - $d_j := 0$, $s_i := s_i - d_j$,
   - $j := j + 1$, go to step 5.
3. If $s_i < d_j$, then
   - $x_{ij} := s_i$, $x_{ik} = 0$ for $k > j$,
   - $s_i := 0$, $d_j := d_j - s_i$,
   - $i := i + 1$, go to step 5.
4. If $s_i = d_j$, then
   - $x_{ij} := s_i$, $x_{ik} = 0$ for $k > j$, $x_{kj} = 0$ for $k > i$,
   - $d_j := 0$, $s_i := 0$,
   - $j := j + 1$, $i := i + 1$, go to step 5.
5. If $i \leq m$ and $j \leq n$, go to step 2.

The idea of this method is to write the solution as an $m \times n$ matrix, and start with the upper-left (or North-West) corner of the matrix. Then, being in ith row and jth column of that matrix, assign the maximum possible shipment between ith source and jth destination, and in the following steps moving down and right. We will refer to this algorithm as the North-West Corner Method.

Let us also notice here, that if all supplies and demand are integral, then a solution of CTP is also integral.

2.2 Modified Transportation Problem

Our next step is to formulate Modified Transportation Problem (MTP). We make the following crucial change in the classical case: transport between some sources and destinations is impossible. We express this condition specifying the set of all possible routes. Moreover, in this paper we are interested only in the existence of a feasible solutions, that is why we omit the objective function. We also assume the problem is already balanced. Thus MTP may be formulated in the following way: given supplies s_i, $i = 1, \ldots, m$, demands d_j, $j = 1, \ldots, n$, and the set $R \subseteq \{1, \ldots, m\} \times \{1, \ldots, n\}$ of possible routes, find any $x_{ij} \in \mathbb{R}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ such that

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = s_i \text{ for } i = 1, \ldots, m, \\ \sum_{i=1}^{m} x_{ij} = d_j \text{ for } j = 1, \ldots, n, \\ x_{ij} \geq 0, \text{ for } (i, j) \in R, \\ x_{ij} = 0, \text{ for } (i, j) \notin R. \end{cases} \tag{MTP}$$

Such a problem will be called an MTP problem with m sources and n destinations. Although we have omitted the goal function, any solution $(x_{ij})_{i=1,\ldots,m}^{j=1,\ldots,n}$ of the above problem will be called a feasible solution of MTP.

The problem formulated above is not new, and can be seen as a special case of a more general Hitchcock Problem [2, 1], but this one is usually stated in terms of the graph theory and solved with combinatorial methods. The problem can be also treated by general linear programming methods, such as simplex [3], however this also complicates the problem, and does not use its special structure. Moreover, these methods do not provide an easy criteria for the existence of the solutions of MTP. In the following we would like to propose such a criterion and an algorithm for finding a feasible solutions of MTP, which essentially makes use of slightly modified North-West Corner Method.

3 Main theorem and algorithm

We are now going to introduce algorithm for finding a feasible solution of MTP, and a criterion for existence of such solutions. When one tries to find such a solution with use of a North-West Corner method, very quickly finds that at some stage, although some sources have positive supply and some destinations have a positive demand, there is no possibility of shipment between these sources and destinations, because the routes between them are not allowed. The idea of our algorithm is to replace the problem with n destinations by a substitute MTP with only 2 destinations, and with the use of obtained solution, to divide the original problem into two MTP problems, both with number of destinations less than n. Applying recursively this procedure we obtain a set of MTP problems with 1 or 2 destinations which can be easily solved.

Let us denote by

$$\overline{D}_j = \{i \in \{1, \ldots, m\} : (i, j) \in R\}, \quad j = 1, \ldots, n,$$

the set of indices of those sources, that are allowed to transport to destination D_j. Let us notice that knowledge of the sets $\overline{D}_j$, $j = 1, \ldots, n$, is equivalent to the knowledge of the whole set R, hence in the formulation fo MTP we could replace the set R with the above sets. Let us now start with the following trivial observation.

Lemma 1. *Modified Transportation Problem with 1 destination have a feasible solution, if*

$$d_1 = \sum_{i \in \overline{D}_1} s_i.$$

Proof. The solution we are looking is given by

$$x_{i1} = \begin{cases} s_i & \text{for } i \in \overline{D}_1, \\ 0 & \text{for } i \notin \overline{D}_1. \end{cases}$$

□

The following lemma is crucial for the algorithm, an shows how to find a solution for a problem with 2 destinations. The idea is to transport from the "exclusive" sources at first, and only then to distribute products from the "shared" sources.

Lemma 2. *Modified Transportation Problem with two destinations have a feasible solution, if*

1. $d_j \leq \sum_{i \in \overline{D}_j} s_i$, $j \in \{1, 2\}$,
2. $d_1 + d_2 = \sum_{i \in \overline{D}_1 \cup \overline{D}_2} s_i$.

Proof. Let us consider the following cases:

1. $d_1 \leq \sum_{i \in \overline{D}_1 \setminus \overline{D}_2} s_i$ and $d_2 \leq \sum_{i \in \overline{D}_2 \setminus \overline{D}_1} s_i$. In this case demands of both destinations are fulfilled by their "exclusive" sources. Thus we can replace this problem with two sub-problems with only one destination that can be solved with the previous lemma: the first MTP with destination D_1 and sources with indices from the set $\overline{D}_1 \setminus \overline{D}_2$, and the second MTP with destination D_2 and sources with indices from the set $\overline{D}_2 \setminus \overline{D}_1$.
2. $d_1 \leq \sum_{i \in \overline{D}_1 \setminus \overline{D}_2} s_i$ and $d_2 > \sum_{i \in \overline{D}_2 \setminus \overline{D}_1} s_i$. Since the sets $\overline{D}_1 \setminus \overline{D}_2$ and $\overline{D}_2$ are disjoint, the demand of the destination D_1 can be fulfilled, by the assumptions of this case, with the supply of its exclusive sources. The demand of the destination D_2, by the assumption $d_2 \leq \sum_{i \in \overline{D}_2} s_i$, may be fulfilled by the supply of sources with indices from the set $\overline{D}_2$. Thus again it suffices to replace this problem with two MTPs with only one destination: first MTP with destination D_1 and sources with indices from the set $\overline{D}_1 \setminus \overline{D}_2$, and the second MTP with destination D_2 and sources with indices from the set $\overline{D}_2$.
3. $d_1 > \sum_{i \in \overline{D}_1 \setminus \overline{D}_2} s_i$ and $d_2 \leq \sum_{i \in \overline{D}_2 \setminus \overline{D}_1} s_i$. This case is symmetric to the previous one.
4. $d_1 > \sum_{i \in \overline{D}_1 \setminus \overline{D}_2} s_i$ and $d_2 > \sum_{i \in \overline{D}_2 \setminus \overline{D}_1} s_i$. This case is the most difficult. In the first step we assign the products to destinations form their "exclusive" sources: entire supply from the sources with indices in the set $\overline{D}_1 \setminus \overline{D}_2$ goes to destination D_1 and whole supply from sources with indices from the set $\overline{D}_2 \setminus \overline{D}_1$ should be transported to D_2. According to the assumptions of this case it is not enough to fulfil their demands. Thus in the next step we form a new, reduced transportation problem, with the same destinations D_1 and D_2, with the sources $\{S_i : i \in \overline{D}_1 \cap \overline{D}_2\}$, and such that demands of the destinations are reduced by the quantities fulfilled in the first step: $\tilde{d}_1 := d_1 - \sum_{i \in \overline{D}_1 \setminus \overline{D}_2} s_i$, $\tilde{d}_2 := d_2 - \sum_{i \in \overline{D}_2 \setminus \overline{D}_1} s_i$. However, the key fact now is that in this problem all the routes are allowed: $R = (\overline{D}_1 \cap \overline{D}_2) \times \{1, 2\}$, since these sources were common to both destinations. By the assumptions of the lemma we have

$$d_1 + d_2 = \sum_{i \in \overline{D}_1 \cup \overline{D}_2} s_i = \sum_{i \in \overline{D}_1 \setminus \overline{D}_2} s_i + \sum_{i \in \overline{D}_2 \setminus \overline{D}_1} s_i + \sum_{i \in \overline{D}_1 \cap \overline{D}_2} s_i,$$

hence

$$\tilde{d}_1 + \tilde{d}_2 = d_1 + d_2 - \sum_{i \in \overline{D}_1 \setminus \overline{D}_2} s_i - \sum_{i \in \overline{D}_2 \setminus \overline{D}_1} s_i = \sum_{i \in \overline{D}_1 \cap \overline{D}_2} s_i.$$

Thus demands of the destinations in the new transportation problem may be fulfilled by the supply of the sources remained in this problem. Since all the routes are allowed, we obtain its solution applying for example classic North-West Corner method. □

We can now pass to the problem with arbitrary number of destinations.

Theorem 1. *Modified Transportation Problem with n destination has a feasible solution if and only if*

$$\bigwedge_{L \subseteq \{1,\ldots,n\}} \left(\sum_{j \in L} d_j \leq \sum_{i \in \bigcup_{j \in L} \overline{D}_j} s_i \right). \tag{1}$$

Proof. Necessity is obvious. We will prove sufficiency by showing how to build a solution by dividing an MTP into sub-problems with a fewer number of destinations. Let us firstly observe that the assumptions of lemmas 1 and 2 are a special cases of the condition (1) for $n = 1$ and $n = 2$, respectively, thus the first steps of induction follows from those lemmas. Let us fix $n > 2$ and assume that the thesis of the theorem is true for all MTP with k destinations, $k < n$.

Let us consider an MTP with n destinations. For non-empty subsets $\emptyset \neq X \subseteq \{1, \ldots, n\}$ we define

$$\Phi(X) := \sum_{i \in \overline{D}_X} s_i - \sum_{j \in X} d_j,$$

where

$$\overline{D}_X := \bigcup_{j \in X} \overline{D}_j,$$

is the set of indices of suppliers that are able to transport their products to destinations with indices form the set X, and let

$$\Phi^* = \min\{\Phi(X) : \emptyset \neq X \subsetneq \{1, \ldots, n\}\}.$$

By the assumption (1) we have $\Phi^* \geq 0$. Let A be any set realizing that minimum:

$$\Phi(A) = \Phi^* \leq \Phi(X), \quad X \subsetneq \{1, \ldots, n\}. \tag{2}$$

Let also $B := \{1, \ldots, n\} \setminus A$. Now we consider the MTP with two destinations, say D_A and D_B with demands equal to $d_A = \sum_{j \in A} d_j$ and $d_B = \sum_{j \in B} d_j$, respectively. The set of sources and their supplies is the same as in the original problem. The set R for this MTP is as follows: $(i, A) \in R \iff i \in \overline{D}_A$, $(i, B) \in R \iff i \in \overline{D}_B$. In other words, we are able to transport products from the source S_i to the combined destination D_A iff we are able to transport the products from the source S_i to any destination with index from the set A. Conditions of lemma 2 are obviously satisfied by this MTP, hence we can use this lemma to obtain a solution: x_{ij}, $i = 1, \ldots, m$, $j \in \{A, B\}$.

Now let us consider two sub-MTPs. First one, say MTP-A, consists of destinations with indices from the set A and all m sources. Demands of the destinations in this MTP are equal to the demands in the original MTP with n destinations: $d_j^A = d_j$, $j \in A$. However, supplies are equal to the values of solution of the MTP with 2 destinations just solved: $s_i^A = x_{iA}$, $i = 1, \ldots, m$. Analogously, second MTP, say MTP-B, consists of destination with indices from the set B and all m sources. Demands and supplies are equal to: $d_j^B = d_j$, $j \in B$, $s_i^B = x_{iB}$, $i = 1, \ldots, m$.

It remains to show that these two sub-MTPs fulfil condition (1). Let us notice that since $s_i^A = s_i$ for $i \in \overline{D}_A \setminus \overline{D}_B$ and $s_i^A = s_i - x_{iB}$ for $i \in \overline{D}_A \cap \overline{D}_B$ we have

$$0 = \sum_{i \in \overline{D}_A} s_i^A - \sum_{j \in A} d_j = \sum_{i \in \overline{D}_A} s_i - \sum_{j \in A} d_j - \sum_{i \in \overline{D}_A \cap \overline{D}_B} x_{iB} = \Phi(A) - \sum_{i \in \overline{D}_A \cap \overline{D}_B} x_{iB},$$

hence $\Phi(A) = \sum_{i \in \overline{D}_A \cap \overline{D}_B} x_{iB}$. Now, for any $Y \subset A$ we have

$$\begin{aligned} \sum_{i \in \overline{D}_Y} s_i^A - \sum_{j \in Y} d_j &= \sum_{i \in \overline{D}_Y} s_i - \sum_{j \in Y} d_j - \sum_{i \in \overline{D}_Y \cap \overline{D}_B} x_{iB} \geq \\ &= \Phi(Y) - \sum_{i \in \overline{D}_A \cap \overline{D}_B} x_{iB} = \Phi(Y) - \Phi(A) \geq 0, \end{aligned}$$

which yields

$$\sum_{i \in \overline{D}_Y} s_i^A \geq \sum_{j \in Y} d_j.$$

For MTP-B we show the condition (1) analogously.

Since MTP-A and MTP-B are problems with less than n destinations and fulfil the condition (1), we infer that they have a feasible solutions, which combined together yields a solution for our original MTP with n destinations. □

4 Example

We now present an example illustrating the above algorithm.

Example 1. Let us consider an MTP with for destinations D_1, D_2, D_3, D_4 and four sources S_1, S_2, S_3, S_4. Table 1 contains the initial conditions of the problem. Numbers in the first column denotes the supplies of the sources, and numbers in the first row are denotes demands of the destinations. Sign "X" in the body of the table denotes that the route between source and destination is not allowed. Cells of the table corresponding to allowed routes are initially empty.

		2	10	2	6
		D_1	D_2	D_3	D_4
4	S_1	X			X
4	S_2				
5	S_3	X		X	
7	S_4		X		

Table 1 Initial problem with 4 destinations.

The difference between supply and demand is lowest for the subset $A = \{D_1, D_2, D_4\}$. Table 2 presents the reduced MTP and its solution. Since $B = \{S_3\}$, the column vector for D_B in the above solution is already a part of

		18	2
		D_A	D_B
4	S_1		
4	S_2		
5	S_3		X
7	S_4		

		18	2
		D_A	D_B
4	S_1	4	0
4	S_2	4	0
5	S_3	5	0
7	S_4	5	2

Table 2 Reduced MTP and its solution.

the final solution. For the set A we have a following sub-MTP. For this problem the difference between supply and demand is lowest for the subset $A = \{D_2, D_4\}$. Table 3 contains also the reduced MTP and its solution. Notice

		2	10	6
		D_1	D_2	D_4
4	S_1	X		X
4	S_2			
5	S_3	X		
5	S_4		X	

		16	2
		D_A	D_B
4	S_1		X
4	S_2		
5	S_3		X
5	S_4		

		16	2
		D_A	D_B
4	S_1	4	0
4	S_2	4	0
5	S_3	5	0
5	S_4	3	2

Table 3 Sub-MTP for 3 destinations, reduced MTP and its solution.

that $B = \{1\}$, hence column of the solution for D_B is a part of the final solution for the destination D_1. Finally,

there remains only one sub-MTP with 2 destination, hence we are able to solve it, and combine with the previous part to obtain a final solution:

		2	10	2	6
		D_1	D_2	D_3	D_4
4	S_1	0	4	0	0
4	S_2	0	4	0	0
5	S_3	0	2	0	3
7	S_4	2	0	2	3

Table 4 A feasible solution.

5 Conclusions

Although the example given above was very simple, we can notice two things. First one is that the main flaw of presented algorithm is the need of computing the function Φ over all subsets of destinations, which has an exponential complexity. However if we are able to find the subset minimizing Φ, the algorithm reduces to less than n very easy problems with at most 2 destinations.

Theorem 1 gives the necessary and sufficient condition for the existence of feasible solutions of a Modified Transportation Problem. To find an optimal solution starting with the feasible one obtained by this method, one may find practical to use the following approach. Use any classical algorithm (eg. method of potentials) with the following alteration: all forbidden routes are now allowed, however a very high penalty cost is assigned to them (higher than the total cost of the initial solution), hence an algorithm will be reluctant to assign any shipments to such routes in the optimal solution.

Surprisingly, a characterization given in this paper have already found an application in cryptography. Namely, with the aid of Theorem 1 one can prove the following result, which we state here without the proof.

Proposition 2. *Let $(J_\ell, \preccurlyeq)$ be a partially ordered set and define the sets:*

$$T = \{(i,j) \in J_\ell \times J_\ell : i \textit{ is a predecessor of } j\},$$

$$H_0^\ell := \left\{ \sum_{(i,j)\in T} \alpha_{ij}(\bar{e}_j - \bar{e}_i) \in \mathbb{Z}^\ell : \alpha_{ij} \in \mathbb{N}_0 \right\}, \qquad H^\ell := H_0^\ell + \mathbb{N}_0^\ell.$$

Let $\bar{w} \in \mathbb{Z}^\ell$. Then $\bar{w} \in H^\ell$ if and only if $\sum_{k\in I} w_k \geq 0$ for every non-empty subset I of J_ℓ with the property

$$\bigwedge_{i\in I} \bigwedge_{j\in J_\ell} (i \preccurlyeq j \Longrightarrow j \in I).$$

References

[1] Dantzig G. B., Thapa M. N.: *Linear Programming 1*, Springer-Verlag New York, 1997.

[2] Korte B., Vygen J.: *Combinatorial Optimization*, 4th Edition, Springer-Verlag Berlin, 2008.

[3] Trzaskalik T.: *Wprowadzenie do badań operacyjnych z komputerem* (in Polish), PWE, Warszawa, 2008.

Three-dimensional Bin Packing Problem with heterogeneous batch constraints.

František Koblasa[1], Miroslav Vavroušek [2], František Manlig[3]

Abstract. The new approaches of Industry 4.0 are focusing more and more on process optimization using principles of artificial intelligence, evolution techniques and computer simulation to achieve both customer and manufacturer goals. One of the biggest issues in this field is to create and optimize production plans. There are lot of well-known solution to a problem of production scheduling of classical homogenous batches, where same types of parts are produced by conventional machines. Nevertheless scheduling of heterogeneous batches is still underestimated, thus not implemented in general planning and scheduling systems.
This article presents possible solution to real world problems where it is necessary to schedule workflow of machines, where it is necessary to take in account its work space while considering several constraints and objective functions. Presented paper is focusing on typical three dimensional bin packing problem, where material of bin assigned parts are influencing total production time and total cost of producing these bin pack batches. Article includes developed constructive algorithm and evolution algorithm which are tested on theoretical examples.

Keywords: Bin Packing Problem, Batch Scheduling, Evolution Algorithm, Constructive Algorithm.

JEL Classification: L23
AMS Classification: 90C27

Introduction

Planning and scheduling becomes one of most important aspects in the Industry 4.0 to meet customer requirements. Nowadays most frequent used planning systems based on ERP (Enterprise Resource Planning system) are focusing on material or capacity planning. APS (Advance Planning and Scheduling) system are focusing to improve this plan by optimizing production batches, job assigning or sequencing to usually find better utilization of manufacturing resources (in most cases machines) or suitable and real due dates. MES systems connected to those mentioned before are than giving us knowledge of actual state of manufacturing together with the option to react on unexpected events.

Those systems are however using classical scheduling approaches of simulating processing of homogeneous batches. Market pressures for greater product variety and forces manufacturing to use smaller and smaller batches which are becoming more diverse. Thanks to new technologies as Rapid prototyping, where we can manufacture different products at once, attention is turning on heterogeneous batches and on the problem of technology compatible and incompatible part families. That leads to necessity to improve ways of planning or better said scheduling in before mentioned information systems.

The attention of this article is put on technologies that used heterogeneous batches processing of one of the oldest technologies in steel industry – heat treating operations especially annealing. The main reason is that, it is very common technology which in the point of view of scheduling has great impact [9,12]. It has possibility to process same and in some cases different part families at one time and it has significant impact on the schedule because its processing time is usually several times greater than the rest of operations, so outcome of optimizing this process has great potential in cost savings.

Scheduling of Annealing furnace is NP hard problem [9], which usually consists of three stages:

1. Constructing job families. This is usually based on material [12], however thickness of the material could also limit placement in the same batch [13].It is possible to process different job families by one processing batch in further described problem.

[1] Department of Manufacturing Systems and Automation, Faculty of Mechanical Engineering, Technical University of Liberec, Studentská 2, Liberec 1, Czech Republic, e-mail: frantisek.koblasa@tul.cz.
[2] Department of Manufacturing Systems and Automation, Faculty of Mechanical Engineering, Technical University of Liberec, Studentská 2, Liberec 1, Czech Republic, e-mail: miroslav.vavrousek@tul.cz.
[3] Department of Manufacturing Systems and Automation, Faculty of Mechanical Engineering, Technical University of Liberec, Studentská 2, Liberec 1, Czech Republic, e-mail: frantisek.manlig@tul.cz.

2. Constructing batches (bins or sacks) by solving Knapsack or Bin packing problem (both usually NP hard problems [12,15]) according to type of the plan[15,16]. It is suitable for short term planning to apply knapsack problem as we are trying to maximize processor utilization or delays of important jobs. However, this article is focusing more on long term planning where we are trying to minimize operation cost (number of batches-bins) taking in to account processing time to meet customer due date as much as possible.
3. Sequencing and batch assigning is the last step in which it is usual to use some of flexible job shop, flow shop or open shop scheduling approach base on the nature of the manufacturing system[7]. It is most probable to use FJSSP [2] in the cases where annealing is in the middle of the process and we don't schedule just furnace, but whole manufacturing system.

Following chapters are dealing with designing batches for Heat treating furnaces. That is made by solving modified three dimensional Bin Packing Problem (3DBPP) where there are given a set of n rectangular-shaped items, each characterized by width wj , height hj and depth dj ($j \in J = \{1;\ldots; n\}$), and an unlimited number of identical three-dimensional containers (bins) having width W, height H and depth D. We assume that the items may be rotated within all axes (x,y,z).

One of the main tasks is to model processing time of heat treating operation together with its objective function. There is presented modified Maximum box constructive algorithm (CA) which results are than compared by designed simple random key based evolution algorithm (EA). For that purpose there are redesigned classical theoretical problems of 3D BPP. Attention is also paid to possible improvements in optimization approach in technology constrains point of view.

Planning 3D BPP batches with constraint of heat treatment

The most difficult aspect of heat treatment scheduling in real world application is determining duration of processing time of batch. There are two basic types of setting batch processing time [3]:

- S-batch – batch processing time is equal to sum of each task in batch.
- P-batch – batch processing time is equal to the maximum processing time among all tasks.

Featured problem of composing manufacturing batch in annealing like production is according before mentioned P-batch, so it is necessary to define processing times of each part to determine maximal. There is big lack of relevant literature which is focusing on determining processing time in annealing operations in gas furnaces (induction furnaces are usually small and with continues process so it is not necessary to solve BPP). This article is trying to combine two major engineering schools (USA and EU) to get at least approximate solution using two types of tools steels (D2 and H13 – US material notation).

USA school [4] is dividing processing time in to three stages of preheating, soaking and tempering time. Data are available for both small parts and also it is possible to determine processing times of bigger parts by incrementing method (see example Table 1).

Steel type	Preheating [min]	Soaking time [min]						Tempering
		< 3 [mm]	< 6[mm]	< 12 [mm]	< 19 [mm]	< 25 [mm]	25 [mm] >	
D2 Tool Steel	15	30	40	50	55	60	60min/25mm	2h/25mm
H13 Tool Steel	10	10	15	20	25	30	30min/25mm	2h/25mm

Table 1: Theoretical time of heat treatment: size-time dependence.

European literature is focusing mostly on smaller parts. There are few new sources but most of them are taking information from Smóling book [13].Smóling in contrast with [4] is considering other aspects of determining processing time – shape of annealed part and distance between parts (see table 2 where a is smallest part dimension - so called characteristic dimension and k_1 and k_4 respective coefficients)

Shape	k_1	Shape	k_1	Placement	k_3	Placement	k_3	Placement	k_3
Sphere or cube	0.75	Short tube	2	Angular or round shape object on pad	1	Several round shape object touching	2	Several angular shape object with gaps 0.5xa	2.2
Round or square block	1	Long tube	4	Round shape on the floor	1	Several angular shape object touching	4	Several roud shape object with gaps 2xa	1.3
Rectangle block 2a xa	1.5	Rectangle block 3-4 a xa	1.75	Angular on the floor	1.4	Several roud shape object with gaps 0.5xa	1.4	Several angular shape object with gaps 1xa	2
Plate	2							Several angular shape object with gaps 2xa	1.8

Table 2: Shape and placement coefficients

This article is than trying to combine those two approaches. There is used approach of summing preheating time soaking time and tempering time and also taking corrective coefficients of shape, placement, type of furnace and material. Processing time (1) is than calculated as

$$t_A = t_{teor} \cdot k_1 \cdot k_2 \cdot k_3 \cdot k_4 \tag{1}$$

Where (t is time and k is corrective coefficient):

- t_A is overall time of heat treatment of part.
- t_{teor} is theoretical time of heat treatment (2) according material and size of characteristic dimension (see Table 1).

$$t_{teor} = t_p + t_s + t_t \tag{2}$$

where t_p is time of preheating, t_s is time of soaking and t_t is time of tempering.

- k_1 is characteristic dimension of part given by Table 2.
- k_2 is influence of heat treating furnace (in our case we are selecting electric powered furnace with k_2=1,2.)
- k_3 is influence of placement in furnace (in our case k_3=4).
- k_4 is division of recommended annealing temperature and recommended quenching temperature. There was selected k_4=0,93 for D2 and 0,79 for H13 tool steel [20].

The goal of following planning optimization is to minimize total competition time of n furnaces-bins of m annealed parts (3)

$$f(x) = \sum_{i=1}^{n} t_{bi} \tag{3}$$

Where t_{bi} (4) is processing time of one batch (annealing furnace)

$$t_{bi} = \max\{t_{A1}; t_{Aj}; \ldots; t_{Am}\} \tag{4}$$

Constructive algorithm to design bins and optimizing evolution algorithm is described in following subchapters.

2.1 Constructive algorithm with completion time for 3D BPP

Constructive algorithm is creating solution in following steps. The set of placed boxes is arranged according to the processing time t_A (1) from the biggest to the smallest. The first block is placed in the corner of the free space. The free space is divided into three new free spaces which covers the space of the cut out boxes. Following boxes are placed in to newly created spaces. Boxes are rotated in all axes if they do not fit at first try. New bin is created if there is no other option of placement. Following blocks (third and following ones) are placed in to newly design free spaces as it is shown at Figure 1. Free spaces which are only subspaces of the bigger ones are eliminated.

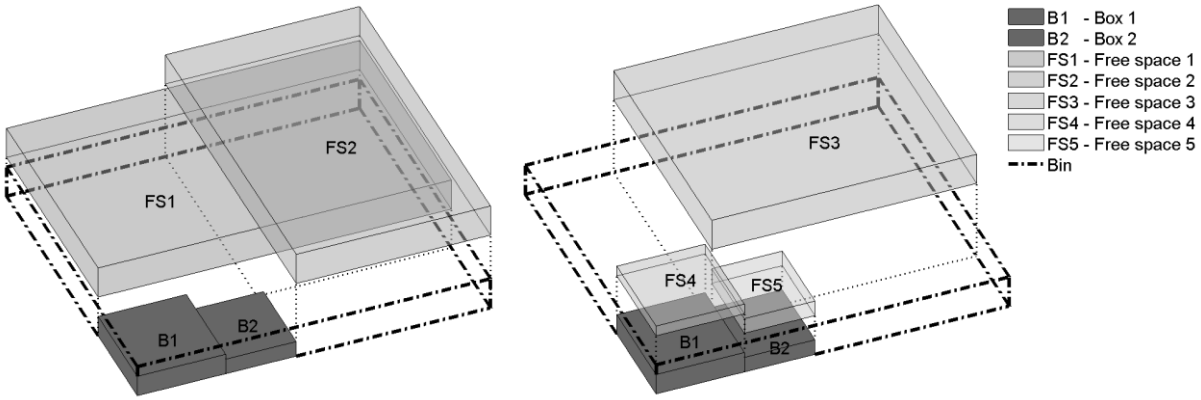


Figure 1: Free space fragmentation

Algorithm is than placing and rotating boxes, by procedure described below, until there are no boxes available:
Data structure
Bins contain set of Bin
Bin contain set of FreeSpace

```
FreeSpace = {w; d; h; centerX; centerY; centerZ}
Boxes contain set of Box
Box = {w; d; h; centerX; centerY; centerZ}
Initialize:
01 create new Bin B in the Bins
02 add new FreeSpace F = (W; D; H) to B
Pack:
03 sort Box in the Boxes by processing time in descending order
04 foreach Box X in the Boxes
05 |   Bool placed = False
06 |
07 |   foreach Bin B in the bins
08 |   |   foreach FreeSpace F in the Bin B
09 |   |   |   for rot = 1:6
10 |   |   |   |   B = rotate X by rot
11 |   |   |   |
12 |   |   |   |   if F.w>=X.w and F.d>=X.d and F.h>=X.h
13 |   |   |   |   |   placed = True
14 |   |   |   |   |   compute F - X and subdivide the result into at most six new FreeSpaces G1;..;G6
15 |   |   |   |   |   add G1;...;G6 to B
16 |   |   |   |   |   break
17 |   |   |   if placed = True
18 |   |   |   |   break
19 |   |   if placed = False
20 |   |   |   create new Bin B in the Bins = (W; D; H)
21 |   |   |   add new FreeSpace F = (W; D; H) to B
22 |   |   |   compute F - X and subdivide the result into at most six new FreeSpaces G1;...;G6
23 |   |   |   add G1;...;G6 to B
24 |   |   optimize FreeSpaces in the Bin B (Erase FreeSpace, which is only subspace)
25 |   |   sort FreeSpaces in the Bin B by size of the volume from the smallest
```

2.2 Random key based evolution algorithm

Beans [2] random key algorithm and its modifications [7]is one of the most versatile evolution algorithm that found its use in wide variety of problems beginning with service[5] or production scheduling [8], traveling salesmen problem[14], container loading problem[20], vehicle routing problem and ending with 2D and 3D bin packing problem [6].

It is suitable for most combinatorial optimization problems whose solutions can be represented by a permutation vector (random key). A random key is a real number, generated at random(uniform distribution) in the continuous interval (0, 1). Decoder is used as a procedure that maps a vector of random keys into a solution of the optimization problem which in our case is sequence of inserted parts in to bins.
Algorithm than has those stages:

1.Generate number of solutions (number of population – *N*) represented by random vectors describing sequence of placing parts in to bins.

2. Calculate fitness function *f(x)* by constructing solutions - see chapter 2.1 equation (2)

3. Sort solutions base on their fitness function in descending order.

4.Select parents – in this case our selecting procedure follows basic random selection. Two parent solutions are selected to breed two new offsprings. Number of selections is equal to size of population *N*

5. Crossover operation to make new offsprings –That ensure all feasible solutions so it is not necessary to repair (or scrap) them as it is usual in knowledge based gene – to gene or one point, two point crossover types.

6. Calculate fitness function *f(x)* of new offsprings.

7. Include offsprings in to new population. Sort new population by *f(x)* by descendent order and let better half to become new generation.

8. Repeat from step 4 until number of defined generations is met.

2.3 Test problems and results

Presented algorithms are tested on modified datasets inspired by 3D BPP class of instances considered by Martello and Vigo in [11]. Instance has bin size W = H = D = 100 and five types of items with two types of material (uniformly random selected) are considered (u.r. stands for „Uniformly random“) with:

- Type 1: w_j u.r. in [1;1/2W], h_j u.r. in [2/3H; H], d_j u.r. in [2/3D; D].
- Type 2: w_j u.r. in [2/3W; W], h_j u.r. in [1;1/2H], d_j u.r. in [2/3D; D].
- Type 3: w_j u.r. in [2/3W; W], h_j u.r. in [2/3H; H], d_j u.r. in [1;1/2D].
- Type 4: w_j u.r. in [1/2W; W], h_j u.r. in [1/2H; H], d_j u.r. in [1/2D; D].
- Type 5: w_j u.r. in [1;1/2W], h_j u.r. in [1;1/2H], d_j u.r. in [1;1/2D].

Our models includes For each type (1-5) there where generated 10 tests with n parts $n \in \langle 10; 15; 20; \ldots; 45 \rangle$. Each model is tested 50 times which means 1750 experiments. Experiments can be divided in two main parts (see Table 3). As there are not known optimal solutions, it was necessary to compare results of CA and EA.

EA algorithm than was set to:

- Size population – its set to double of size of problem (n items) + 100 to ensure rich gene pool for optimization
- Crossover – uniform crossover is used with probability 80% to exchange genes.
- Parent selection – in compare to [6] it is used uniform selection.
- New generation selection – it is used classical elitism approach, so after merging both parent and offspring population, the better half is than selected to survive in to new generation.
- Number of generation – it usual to use at least 100 generation to improve solution properly, however its used just 20 generation because of extensive optimization time.

The first part of the table shows how EA was able to reduce fitness function generated by CA (percentage showed EA fitness function divided by CAs). Second part of the table (time gain) shows time in minutes which we can get by optimizing solution of CA by EA. That means solution improvement reduced by optimization timespan.

	Fitness function reduction down to					Time gain by optimization [min]				
n. items/model	1	2	3	4	5	1	2	3	4	5
10	94.79%	70.07%	64.53%	97.39%	75.08%	108.56	341.28	381.50	91.84	126.31
15	91.01%	93.79%	92.15%	91.91%	73.02%	173.86	107.81	182.98	397.09	399.07
20	93.85%	86.97%	83.22%	93.28%	78.46%	140.22	302.50	288.73	463.53	318.80
25	92.39%	100.14%	95.29%	93.70%	80.06%	225.46	-34.38	97.51	497.40	296.11
30	91.37%	94.96%	94.67%	94.85%	80.65%	261.99	185.78	198.38	444.99	594.41
35	92.55%	97.94%	100.01%	100.00%	77.73%	275.76	2.07	-76.81	-68.23	505.16
40	101.82%	97.41%	96.18%	99.16%	90.52%	-183.17	60.70	126.74	64.71	260.37

Table 3: Optimization results of CA and EA

Preliminary test of designed EA and CA algorithm shows that in some cases EA algorithm is able to reduce value of fitness function down to 64% of CA value (model 3 with 10 items). However there were models that EA can't optimize to better solution than CA, in fact solutions were in average worse (25 items model 2, 35 items models 3 and 4 and 40 items model one). On other hand, possible time gained thanks to the optimization is in most cases and in best case EA was able to save nearly 10 hours (problem 5 with 30 items to pack) The main problem lies in very long optimization time which limits number of generations during EA optimization where CA timespan takes less than seconds and EA at least more than 3 minutes up to hour and half which could be complication in the practice application if you need plan just in time.

Conclusion

Designed simple random key evolution algorithm shows in most cases promising results in comparison of constructive algorithm thus i tis necessary to focus not only on objective (fitness) function but also on the optimization timespan. Despite the fact that number of generations in EA was set low (20) optimization timespan was high. The main reason is searching pattern in constructive algorithm which is actively looking for optimal rotation in particular constructing step. That makes part of CA more like local search and it is extending constructing solution significantly. Further work in this field will focus on adaptive setting of population size a number of generations base on quality of fitness function and also on representation of rotation problem by chromosome.

This article also presented new way how to set processing time of batch scheduling in heat-treating operations like annealing. Although this approach is following verified methods, it is combining principally completely different material engineering schools, so this approach can't be used yet for practice in the terms of material engineering without deeper material and technology knowledge. D2 and H13 were also selected as compatible families only because their processing times are significantly different - so interesting in the point of combinatorial optimization view. That does not mean they could be combine in the process of annealing.

Nevertheless, the way to calculate processing time is principally feasible and preliminary results of processing times were consulted with field professionals and were considered as mostly acceptable. This new approach will be further consulted with material, technology and also planning and scheduling experts. Further research in this field would be focusing also on how gaps between placed objects can influence total production time (see Table 2 influence of k_3 coefficient).

Acknowledgements

The research reported in this paper was supported by targeted support for specific university research within the student grant competition TUL (Project 21130) - Research and development in the field of 3D technology, manufacturing systems and automation.

References

[1] Baldi, M. M., Perboli, G., and Tadei, R.: The three-dimensional knapsack problem with balancing constraints. *Applied Mathematics and Computation*, **218(19)**, (2012),9802-9818

[2] Bean, J. C.: Genetic algorithms and random keys for sequencing and optimization. *ORSA journal on computing*, **6**, no.2, (Spring 1994)

[3] Brucker, P. *Scheduling Algorithms*. 5th ed. Berlin: Springer, 2007, 371 s. ISBN 978-3-540-69515-8.

[4] Bryson, W. E.: *Heat Treatment Selection and Application of Tool Steels 2E*. Hanser Verlag, Munich, (2005), ISBN 978-3-446-41941-4

[5] Damm, R. B.; Resende, M. G.C nad Ronconi, D. P.: A biased random key genetic algorithm for the field technician scheduling problem. *Computers & Operations Research*, 2016, **75**: 49-63.

[6] Gonçalves, J. F., and Resende, M. G.: A biased random key genetic algorithm for 2D and 3D bin packing problems. *International Journal of Production Economics*, **145(2)**, 2013, 500-510.

[7] Gonçalves, J. F.and Resende, M.G.C.: Biased random-key genetic algorithms for combinatorial optimization. *Journal of Heuristics*, **17.5**: (2011), 487-525.

[8] Gonçalves, J. F.and Resende, M.G.C.:. An extended Akers graphical method with a biased random-key genetic algorithm for job-shop scheduling. *International Transactions in Operational Research*, **21.2**, (2014), 215-246.

[9] Jacobi, S., León-Soto, E., Madrigal-Mora, C., and Fischer, K.: MasDISPO_xt: Annealing Furnace Planning inside the Supply Chain of Steel Production. *IFAC Proceedings* , **41(2),** (2008), 13888-13892.

[10] Li, G., Li, X., & Wu, J. Study of the thermal fatigue crack initial life of H13 and H21 steels. *Journal of materials processing technology*, **74(1)**, (1998), 23-26.

[11] Martello, S., Pisinger, D., and Vigo, D.: The three-dimensional bin packing problem. *Operations Research*, **vol 48**, 2, (2000), 256-267.

[12] Mathirajan, M., Chandru, V., and Sivakumar, A. I.: Heuristic algorithms for scheduling heat-treatment furnaces of steel casting industries. *Sadhana*, **32(5),** (2007), 479-500

[13] Smóling K.: *Tepelné a chemicko-tepelné spracovanie v príkladoch*. Alfa, Bratislava,1988.

[14] Snyder, L. V.; Daskin, M. S: A random-key genetic algorithm for the generalized traveling salesman problem. *European Journal of Operational Research*, **174.1**,(2006), 38-53.

[15] Stawowy, A. and Duda, J.: Models and algorithms for production planning and scheduling in foundries – current state and development perspectives. *Archives of Foundry Engineering*. **12(2)**, (2012) 69-74. DOI: 10.2478/v10266-012-0039-4

[16] Szőke, Á.: Bin-packing-based planning of agile releases. In: *International Conference on Evaluation of Novel Approaches to Software Engineering*. Springer Berlin Heidelberg. (2008, May). 133-146

[17] Večeřová, V. *Tepelné zpracování – 4.R. - cvičení: Strojírenská technologie*. Ostrava, 2012

[18] Večeřová, V. *Tepelné zpracování – 3.R. - cvičení: Strojírenská technologie*. Ostrava, 2012

[19] Xie, X., Li, Y., & Zheng, Y.: Multiple crane scheduling in batch annealing process with no-delay constraints for machine unloading. In: *International Conference on Information and Automation (ICIA)*, (2012). 597-601

[20] Zheng, J.-N.; Chien, Ch.-F.; Gen, M.: Multi-objective multi-population biased random-key genetic algorithm for the 3-D container loading problem. *Computers & Industrial Engineering*, **89** 2015, 80-87.

Spatial Modelling of Nominal and Real Net Disposable Household Incomes in the Districts (LAU 1) of the Czech Republic

Aleš Kocourek[1], Jana Šimanová[2]

Abstract. The aim of this article is to suggest and apply spatial econometric model for estimation of the regional net disposable household incomes at the district level. Its purpose is to provide an instrument for more precise and more realistic comparison of the income level of average households across the regions of the Czech Republic in nominal value and also in purchasing power parity (real value).
The article extends the application of the results of the research project devoted to issues of regional price levels in the Czech Republic. Therefore, it also contributes to solution of the often-discussed problem of nominal vs. real income indicators as benchmark of socioeconomic disparities.
Authors use statistical method of Small Area Estimation involving the estimation of parameters for small sub-populations, generally used when the sub-population of interest is included in a larger survey. Estimated nominal and real values of the regional net disposable income of average household are statistically analyzed. The findings underpin the need for a more accurate specification of economic and social disparities on a lower regional level.

Keywords: Net household disposable income, Regional price level, Small area estimation, LAU 1.

JEL Classification: C23, R13
AMS Classification: 62P20, 91B72

1 Introduction

The aim of this article is to suggest and apply spatial econometric model for estimation of the regional net disposable household incomes at the district (Local Administrative Units, LAU 1) level. Its purpose is to provide an instrument for more precise and more realistic comparison of the income level of average households across the regions of the Czech Republic in nominal value and also in purchasing power parity (real value). The article continues and extends application of the results of the research project devoted to issues of regional price levels in the Czech Republic. [12][14] Therefore, it also contributes to solution of the widely-discussed problem of nominal vs. real income indicators as benchmark of socioeconomic disparities [2].

The fundamental research hypothesis claims, the higher levels of income of households (measured by the nominal net disposable household income, *NDHI*) generally tend to be compensated for by higher levels of price levels. Therefore, the comparison of nominal values of *NDHI* across regions does not illustrate the real socioeconomic position of the region's inhabitants, although the nominal *NDHI* is generally used as the measure of socioeconomic ranking of regions in the Czech Republic and other European countries. In the strategy of the regional development of the Czech Republic, however, the nominal net disposable income is one of the crucial indicators determining the social position of inhabitants of a region.

The quantification and evaluation of regional disparities remains one of the most up-to-date topics of regional politics. According to Czech as well as foreign authors, the role of the supply side is often overestimated in the regional policy at the expense of the demand side, or more specifically of the real income per capita. The effect of the level of regional living costs or regional price levels is perceived by the current theories of regional development as an impact of localization of corporations. It is presumed (to a great extent controversially) that the consumer price levels are lower and the real estate price levels are higher as a result of economies of agglomeration [19]. According to Viturka [21], the price factors belong to the group of moderately

[1] Technical University in Liberec/Faculty of Economics, Department of Economics, Studentská 1402/2, 460 01, Liberec, ales.kocourek@tul.cz.
[2] Technical University in Liberec/Faculty of Economics, Department of Economics, Studentská 1402/2, 460 01, Liberec, jana.simanova@tul.cz.

important determinants of regional competitiveness. Kahoun [10] considers the fact that the regional differences in price levels remain neglected highly limiting for accountable regional comparison, especially because the difference in price levels between the Czech regions are significant. Differences in the price levels across the ČR at the LAU 1 were proved by Kocourek, Šimanová [12], these results as the original and unique on the territorial level LAU 1 will be used in this article.

Since the value of *NDHI* is not published on disaggregated to the level of district cities in the Czech Republic, but only to the level of regions (NUTS 3), it was necessary to suggest and apply spatial econometric model for estimation of the regional net disposable household incomes at LAU 1 and therefore select potential regressors on the macroregional (NUTS 3) level available also for the districts (LAU 1). Authors use statistical method of Small Area Estimation involving the estimation of parameters for small subpopulations, generally used when the sub-population of interest is included in a larger survey. These subpopulations should be characterized as smaller areas (districts). [15] Estimated nominal and real values of the regional net disposable income of average households are statistically analyzed. The findings underpin the need for a more accurate specification of economic and social disparities on a lower regional level.

Previous research papers document that in the German NUTS 3 regions the real regional disparities were proved to diminish at a higher pace than the nominal ones, especially across the East German regions [13][18]. In the United Kingdom, the issue of real regional disparities has been tackled by Overman and Gibbons [7], who focus solely on the prices of housing. During their research in 1998–2008, a significant trade-off between the level of wages and the costs of living was identified. Therefore, they recommend the economic policies should target the individual inhabitant and should attempt to improve his/her individual position, which will result in raising the situation of the whole region more efficiently than focusing on a geographically determined region [7]. In the USA, the researchers from the Bureau of Economic Analysis are deeply engaged in the issue of metropolitan and nonmetropolitan price indices and also in the context of real income of population. They discovered a higher variability in real incomes in the nonmetropolitan areas than in the metropolitan ones. [1]

2 Methodology of Estimation of the Net Disposable Household Income at the District (LAU 1) Level

The net disposable household income (*NDHI*) is the result of balance of revenues and expenditures recorded at the secondary distribution of income account. It shows how is the surplus/deficit of primary incomes redistributed through the taxes, social benefits and other transfer payments [16]. The *NDHI* as such represents a value of money the households are at their disposal for final consumption, savings, or assets accumulation. The indicator illustrates the material wealth of households with permanent residence in the particular region or locality [7].

Since the value of *NDHI* is not published on disaggregated to the level of district cities in the Czech Republic, but only to the level of regions (NUTS 3), it was necessary to select potential regressors on the regional (NUTS 3) level available also for the districts (LAU 1) of the Czech Republic to estimate the income indicator *NDHI* per capita in the requested territorial structure and time. A group of potential regressors, whose influence on the *NDHI* seemed logically justifiable and statistically significant, was chosen and tested. For the estimation of model, the panel data analysis methods were applied using the software eViews 8.1.

The aim of the estimation is to regionally disaggregate the data (not to perform forecasting or extrapolation in time). We ran the augmented Dickey-Fuller test (ADF) to test for the stationarity of the data. It indicates the data are geographically stationary. Regarding the characteristics of the data, fixed effects were chosen for the period and for the cross-section dimension. The results of the testing are summed up in the model equation (1) bellow and in the figure 1 showing the model's performance on the actual data on the regional (NUTS 3) level. The fixed cross-section and period effects were estimated using ordinary least squares.

$$NDHI_t^r = 264{,}170.515 - 5{,}144.441 \times PRIMEDU_t^r + 2{,}884.807 \times UNIEDU_t^r - 1{,}781.867 \times UNEMP_t^r \qquad (1)$$

where *NDHI* is the nominal value of net disposable household income, *PRIMEDU* is the share of population at the age of 15+ that has attained pre-primary or primary level of education only, *UNIEDU* is the share of population at the age of 15+ that has attained the tertiary level of education, *UNEMP* is the rate of unemployment (the share of unemployed persons on the labor force), *r* is the region (here NUTS 3), and *t* is the year (2005–2015).

Other regressors available at the LAU 1 as well as at the NUTS 3 level (such as share of employees in the primary sector on the economically active population, share of employees in the secondary sector on the economically active population, specific population density, average market price of a dwelling, share of

population at the age from 15 to 60 years, share of population living in the district city) we tested and proved to be statistically insignificant.

The statistically significant regressors of *NDHI* on the level of regions (NUTS 3) included *PRIMEDU* and *UNEMP* with negative influence on the *NDHI* and *UNIEDU* with a positive influence on the *NDHI*. These three variables are capable of explaining 91.01 % of variability of the data, the robustness of the model was also verified using *F*-statistics and Jarque-Bera test for normality of residuals.

We applied the equation (1). We assigned the estimated fixed cross-section effects to all the districts (LAU 1) of a certain region (NUT 3) and calculated the values of *NDHI* with the data (*PRIMEDU, UNIEDU, UNEMP*) characterizing each individual district of the Czech Republic in each year of the period 2011–2015. Thus, we received the preliminary *NDHI* estimates showing the income differentiation within each of the regions (NUTS 3) of the Czech Republic. We then rectified all the *NDHI* values so that their regional average weighted by the population of each district corresponds with the real *NDHI* value at NUTS 3 level [5].

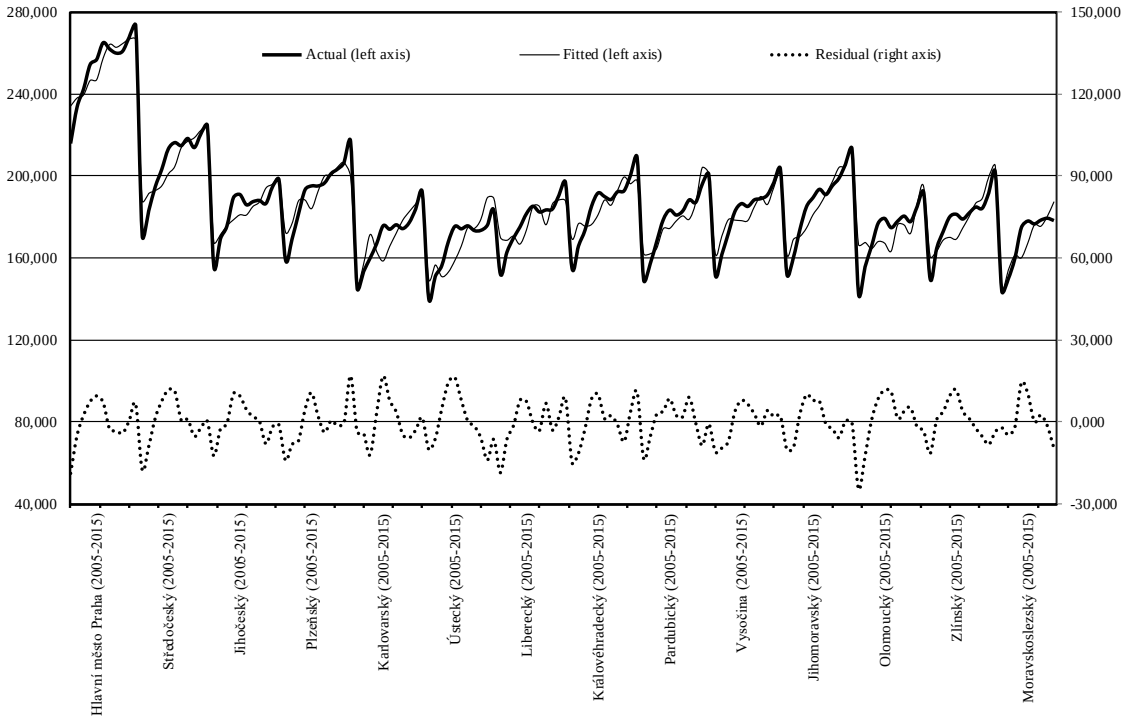


Figure 1: Fitted, Actual and Residual Values of the Regression Model of Regional Nominal NDHI in the Czech NUTS 3 Regions in 2005–2015
Source: authors' own calculations in eViews 8.1 based on the data from [6]

2.1 Adjusting the Regional Net Disposable Household Incomes through Regional Price Levels

The software eViews 8.1 was employed to estimate the values of *NDHI* in the "lower" regional decomposition of districts (LAU 1). The results can be found in the figure 2, illustrated as the nominal *NDHI* values for the 76 districts in the years 2011–2015.

The resulting nominal *NDHI* estimates at the district level (LAU 1) of spatial decomposition are further adjusted by the value of the regional multilateral transitory price level index – the Regional Price Index (*RPI*). The *RPI* was methodically designed and calculated in the research project "Regional Price Index project as an indicator of real and social economic disparities". [14] Index values for 76 LAU 1 are available in the public database [22] and were published in the peer-reviewed statistical journal [12]. The regional price-level adjusted real net disposable household income (*rNDHI*) was calculated according to the formula:

$$rNDHI_t^r = \frac{NDHI_t^r}{RPI^r \times CPI_t}, \tag{2}$$

where *NDHI* is the nominal value of net disposable household income, *r* is the region (here the district LAU 1), *RPI* is the regional price index, *CPI* is the national consumer price index, and *t* is the year (2011–2015).

3 Results

The nominal and real *NDHI* were calculated for the 76 districts in years 2011–2015. The results are shown in Figures 2. The *rNDHI* is very useful for mapping of the development of the socioeconomic position of an average household residing in a particular district of the Czech Republic over the period of 2011–2015. As expected, the most pronounced differences are recorded in the largest cities of Praha, Brno, and perhaps Plzeň. Figures 2 and 3 as well as descriptive statistics shown in Table 1 suggest, there might be some trade-off between the regional values of nominal *NDHI* per capita and the regional price levels of *RPI*: the higher nominal income (*NDHI*) seems to be compensated for by the higher values of regional price levels (*RPI*).

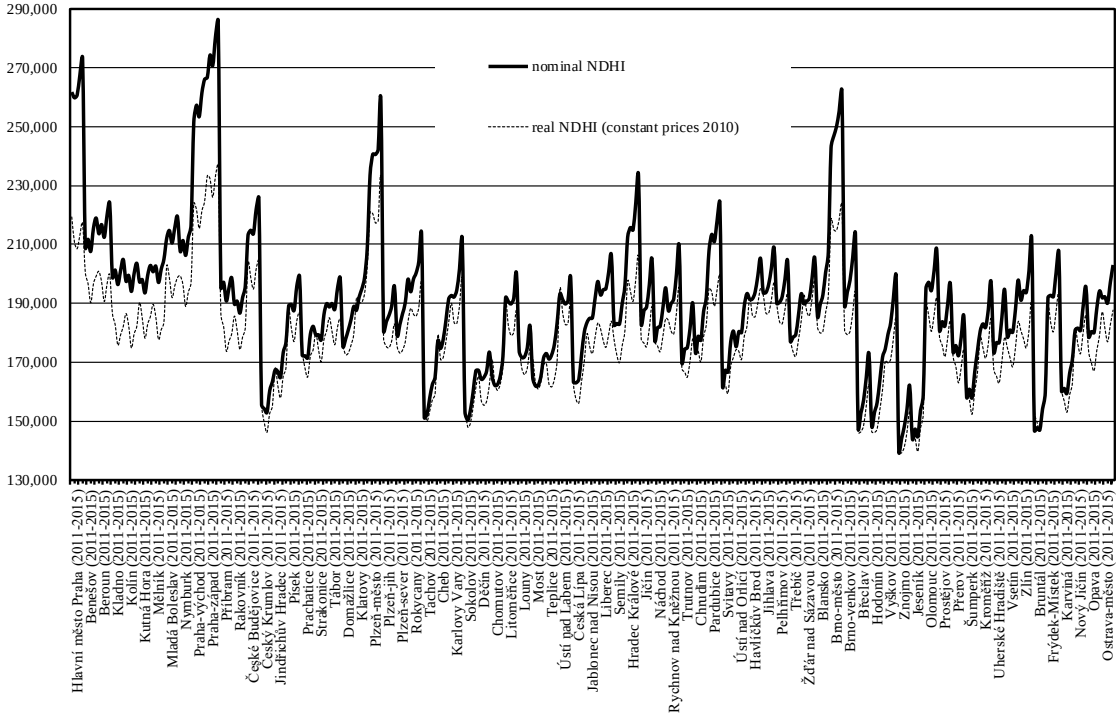


Figure 2: Nominal NDHI vs. Real NDHI in the Czech LAU 1 Districts in 2011–2015
Source: authors' own calculations in eVeiws 8.1 based on the data from [6]

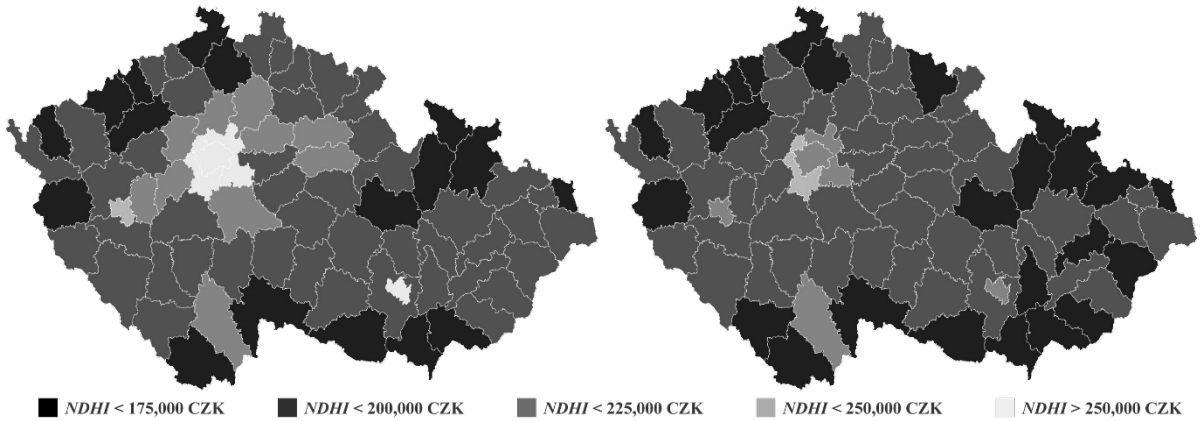


Figure 3: Nominal NDHI (left) vs. Real NDHI (right) in the Czech LAU 1 Districts (2011–2015 Average)
Source: authors' own calculations in eVeiws 8.1 based on the data from [6]

Indicator	Standard Deviation	Mean	Median
nominal *NDHI*	25,125.004	190,729.905	189,888.630
real *NDHI*	17,502.134	179,427.196	179,146.683

Table 1: Descriptive Statistics of the Nominal and Real NDHI
Source: authors' calculations in eVeiws 8.1 based on the data from [6]

The main hypothesis of this article was focused on validation of the statistically significant influence of regional price levels on the extent of recorded interregional nominal socioeconomic disparities. We used five tests on homogeneity of variance (F-test, Siegel-Tukey test, Bartlett test, Levene test, and Brown-Forsythe test). The results of all these tests are summed up in the Table 2. All of them give very similar answers.

Since the *P*-value is lower than the 5% significance level, the null hypothesis of homogeneity of variances is rejected. We can conclude the regional price levels reassess the nominal regional disparities significantly at the 95% confidence level. The interregional differences measured by nominal *NDHI* are significantly wider than the real *NDHI* disparities.

Method	Degrees of Freedom	Value	*P*-value
F-test	(384, 384)	2,061	0,000
Siegel-Tukey		3,858	0,000
Bartlett	1	49,073	0,000
Levene	(1, 768)	20,193	0,000
Brown-Forsythe	(1, 768)	19,915	0,000

Table 2: Results of the Tests on Homoscedasticity
Source: authors' calculations in eViews 8.1 based on the data from [6]

4 Conclusion

The regression analysis on panel data showed that *NDHI* per capita in the regions of the Czech Republic is influenced mainly by the rate of unemployment and the level of education of the inhabitants. The results of analysis across all examined districts of the Czech Republic verified the statistically significant trade-off between the price levels and *NDHI*: the higher incomes generally imply higher regional price levels. The analysis of the variability of the statistical set of nominal *NDHI*s per capita was tested against the variability of the real *NDHI*s per capita (in constant prices of 2010). The significant impact of the regional price levels was verified at the 95% confidence level. The nominal indicator of socioeconomic position (*NDHI*) of an average individual in the 76 districts of the Czech Republic recorded significantly higher variability than real indicator (*rNDHI*). Thus, the differences in prices across regions decrease the interregional disparities and to a certain extent improve the socioeconomic situation of inhabitants especially in the problematic, structurally affected regions of the Czech Republic.

Although the differences in the quality of services such as education or health care are not accounted for, the spatial assessment of the multilateral regional price differences has the potential of improving the understanding of some of the market inefficiencies and represents an important mean of more precisely targeted interventions of economic policy [20]. The regional price levels play a crucial role in consumers' decision making, in localization of economic subjects, and as such can widely affect the regional disparities [2][8].

More precise definition of localities as well as methods of assessing the real economic and social disparities is desirable for increasing the efficiency of applied instruments of regional policies [11]. It seems useful to focus the policies of regional development on the real socioeconomic situation of the individuals. The real income indicators make the analysis of social and economic disparities on the regional and sub-regional levels more precise [21][10]. The real regional disparities in the income of an average household across the Czech Republic are smaller than so far published nominal ones, which is consistent with findings of [3] or [17].

Therefore, it seems very useful (if not necessary) to measure or at least estimate the price levels on the most detailed scale available. Significant differences in cost of living can be identified even within the former districts in the Czech Republic (LAU 1). From this point of view, a price level homogeneity on the level of NUTS 3 or even NUTS 2 is a very strong and hardly justifiable precondition.

Acknowledgements

This research is a part of the applied research project TD020047 "Regional Cost-of-Living Index as the Indicator of the Real Social and Economic Disparities" supported by the Technology Agency of the Czech Republic, the Omega Programme. The article was prepared with an institutional support of the long-term conceptual development of the Faculty of Economics, Technical University of Liberec, in the framework of the project Excellent Research Teams – Regional Development of the Czech Republic in the Context of the Onset of the Fourth Industrial Revolution.

Our special thanks belong to the staff of the Price Statistics Department of the Czech Statistical Office for their willingness to provide data and other information related to data collection and processing.

References

[1] Aten, B., Figueroa, E., and Martin, T. *Real Personal Income and Regional Price Parities for States and Metropolitan Areas, 2007–2011* [online]. Washington, USA: U.S. Bureau of Economic Analysis, 2013. [cit. 2014-04-06]. Available from WWW: http://www.bea.gov/scb/pdf/2013/08 August/0813_regional _price_parities.pdf

[2] Brandová, B. Regional Economic Growth. In *Proceedings of the 10th International Conference Liberec Economic Forum 2011* (Kocourek, A. ed.). Liberec: Technical University of Liberec, 2011, 34–42.

[3] Čadil, J., Mazouch, P., Musil, P., and Kramulová, J. True Regional Purchasing Power: Evidence from the Czech Republic. *Post-Communist Economies*, 2014, **26**(2), 241–256.

[4] Čajka, P. Regionálna politika v kontinentálnom kontexte. In *Rozširovanie Európskej únie na juhovýchod a Slovenská republika* (Kosír, I. et al. eds.) Bratislava: Ekonóm, 2007, 14–50.

[5] Chandra, H., Salvati, N., Chambers, R., and Tzavidis, N. Small Area Estimation under Spatial Nonstationarity. *Computational Statistics and Data Analysis*, 2012, **56**(10), 2875–2888.

[6] CzSO. *Data from Internal Database of Czech Statistical Office.* Praha: Czech Statistical Office, 2014.

[7] CzSO. *Indexy spotřebitelských cen (životních nákladů) – základní členění únor 2015* [online]. Praha: Czech Statistical Office, 2015. [cit. 2015-03-11]. Available from WWW: http://www.czso.cz/csu/2015 edicniplan.nsf/t/000038003F/$File/012018-150206.pdf

[8] Dědková, J. and Skrbková, D. Significant Competitiveness Factors of Companies in the Czech Part of the Euroregion Neisse-Nisa-Nysa. In *Proceedings of the 8th European Conference on Innovation and Enterpreneurship* (Teirlinck, P., Kelchtermans, S., and DeBeule, F. eds.). Bruxelles: Hogeschool-Universiteit Brussel, 2013, s. 212 – 220. ISBN 978-1-909507-62-3.

[9] Gibbons, S., Overman, H., and Resende, G. Real Earnings Disparities in Britain. *Spatial Economics Research Centre Discussion Paper*, **2011**(65), 44 pgs.

[10] Kahoun, J. Měření regionálního HDP: důchodový a produkční přístup. *Ekonomické listy*, 2011, **2**(5), 3–13.

[11] Kocourek, A., Šimanová, J., and Kraft, J.: Regionalization of the Consumer Cost-of-Living Index in the Czech Republic. In *The 8th International Days of Statistics and Economics Conference Proceedings* (Löster, T. and Pavelka, T. eds.). Praha: University of Economics, 2014, 1487–1496.

[12] Kocourek, A., Šimanová, J., and Šmída, J.: Estimation of Regional Price Levels in the Districts of the Czech Republic. *Statistika: Statistics and Economy Journal*, 2016, **96**(4), 56–70.

[13] Kosfeld, R., and. Eckey, H.: Market access, regional price level and wage disparities: the German Case. *Jahrbuch für Regionalwissenschaft*, 2010, **30**(2), 105–128.

[14] Kraft, J. et al.: *Regionální cenové hladiny – teorie, metodika a praxe*. Technical university of Liberec, Liberec, 2015.

[15] Longford, N. T.: *Missing Data and Small Area Estimation: Modern Analytical Equipment for the Survey Statistician*. Springer Science & Business Media, Berlin, 2006.

[16] Maršíková, K., and Kocourek, A.: Shifts in Income Expectations of Czech Students at Selected Economic Faculties over the Years 2001–2012. *Ekonomický časopis*, 2013, **61**(4), 358–375.

[17] Pavelka, T., Skála, M., and Čadil, J.: Selected Issues of the Minimum Wage in the Czech Republic. *E+M Economics & Management*, 2014, **17**(4), 30–45.

[18] Roos, M. W.: Earnings Disparities in Unified Germany: Nominal versus Real. *Jahrbuch für Regionalwissenschaft*, 2006, **26**(2), 171–189.

[19] Schultze, C. L.: The Consumer Cost-of-Living Index: Conceptual Issues and Practical Suggestions. *Journal of Economic Perspectives*, 2003, **17**(1), 3–22.

[20] Šimanová, J. and Trešl, F.: Vývoj průmyslové koncentrace a specializace v regionech NUTS 3 České republiky v kontextu dynamizace regionální komparativní výhody. *E+M Economics & Management*, 2011, **14**(1), 38–52.

[21] Viturka, M.: Konkurenceschopnost regionů a možnosti jejího hodnocení. *Politická ekonomie*, 2007, **55**(5), 638–658.

[22] Žižka, T., Kocourek, A., and Šimanová, J.: *Strukturovaná výpočtová databáze regionálních cenových indexů pro měření regionálních cenových hladin v členění dle CZ-COICOP* [software]. Available from WWW: http://vyzkum.ef.tul.cz/td020047/index.php?content=databaze_new

Consumption taxes in Australia: A DSGE approach

Zlatica Konôpková[1]

Abstract. This paper investigates the effects of the increase in consumption taxes on Australian economy which faces the same challenges as almost all developed countries of the world. The public finances are in weak condition recovering from the last global financial crisis. Current status of demography prognoses the future decrease of working population and significant increase in the costs on health and social system. The goal is to decide whether currently discussed need of tax reform is relevant and if the proposed solution can improve the situation of public finances using a new Keynesian dynamic stochastic general equilibrium model for a small closed economy which is adjusted for the changes in consumption taxes. The effects on economy (especially on output, deficit and debt) are examined using impulse response functions of shocks in the stochastic components of the implicit tax rate. According to our results we do not suggest Australian government to implement the reform of increasing indirect taxes itself, but combine it with e.g. decrease of direct taxes as it is the main trend in fiscal policy of nowadays.

Keywords: DSGE, consumption taxes, tax reform, Australia.

JEL classification: H21, C50
AMS classification: 91B64

1 Introduction

The latest global financial crisis tested the condition of government budgets and most of them have not outperformed the consequences yet. The public finances are in weak condition, slowly recovering from the crisis. In addition, developed countries face the problem of aging population indicating the future decrease in working population associated with significant increase in the costs on health and social system. Australia is not an exception and deals with the same problems.

In the following years, the consolidation of public sector is a crucial task for policy makers. They must find the equilibrium in which they collect sufficient amount of funds to cover their activities and do not influence the private sector too much at the same time. Taxes may have negative impact on economy. They burden the part of economic activity, create distortions and may slow down the economic growth.

The aim of this article is to analyze possible effects of increasing the taxes on consumption on Australian economy using dynamic stochastic general equilibrium (DSGE) model. In the following chapters, we present the details of current situation of Australia and the details of the chosen DSGE model. The original model is adjusted in order to simulate the shock in consumption taxes and then calibrated. The effects of the shock are discussed using the impulse response functions.

2 Current problems

The consequences of the last global financial crisis are still visible in economies. Australian budget has fallen from surplus into deficit $5.61\ \%$, currently accounting for deficit $2.8\ \%$ (OECD [9]). The debt has almost doubled since 2007 and reached the level $64.2\ \%$ of GDP.

In addition, Australia faces the problem of aging population predicting the pressure on health and social system. Figure 1 shows the structure of Australian population and changes in the decade between 2004 and 2014. The number of people aged 65 and over increased in last decade by 1.87 percentage point. On the other side the youth population decreased by 1.11 percentage point and there are no signals of changing the trend in the near future (OECD [8]).

It is inevitable to consolidate the public finances and the pressure is put on policy makers to prepare the effective fiscal reforms. They can focus on cutting their spending or increasing the taxes. The distortion effect of the taxes is different for all types and therefore the structure of tax system can affect the success of taxes collection. According to OECD [12] consumption taxes and property tax have the smallest negative effects on economy and the corporate and personal income taxes are on the opposite side. Nowadays the main trend in fiscal policy is to make a move

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno, Czech Republic, konopkova.zlatica@mail.muni.cz.

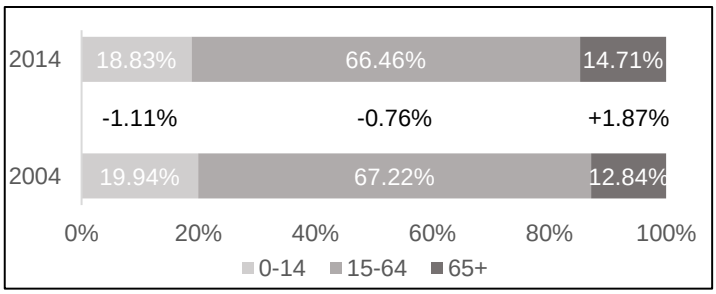


Figure 1 Changes in the structure of Australian population 2004 – 2014 (OECD [8])

from direct taxes to indirect ones. Mamatzakis [7] confirmed that this kind of shift can invoke an economic growth without changes on revenue side.

As it is shown by Figure 2 the direct taxes compound more than 2/3 of Australian taxation and therefore there is a place for reform in favor of indirect taxes which is nowadays discussed and highly promoted by institutions in Australia in order to gain the public's vote.

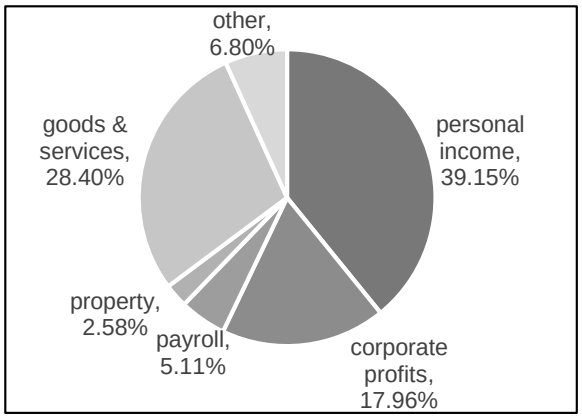


Figure 2 Taxes in Australia – % of taxation (OECD [11])

3 Methodology and data

The model chosen for analysis is an adjusted version of model presented in Costa Junior and Sampaio [4] who examined the tax reduction effects of the productive sector in Brazilian economy. The model simulates the environment of small closed economy divided into three sectors: households, firms and government (authority). The omission of international trade is motivated by the intention to use as much simple model as possible and also by the fact that Australia is the 3rd least open economy in OECD with trade only 41 % of GDP (OECD [14]).

Households

The whole sector is divided into two types of representative agents – *active* (Ricardian) and *inactive* (non-Ricardian) workers. The split of sector is fixed, thus the aggregate consumption C_t of individuals is equal to

$$C_t = (1-\omega)C_{R,t} + \omega C_{NR,t}, \tag{1}$$

where $C_{R,t}, C_{NR,t}$ marks the consumption of Ricardian (non-Ricardian) households and ω the ratio of non-Ricardian households in economy.

Active workers contribute to the pension system, pay taxes and also make savings. They choose the level of consumption, savings, investments and leisure in order to maximize their intertemporal utility function. Let E_t be the expectation operator, $\beta \in (0,1)$ the intertemporal discount factor, L_t the labor, ψ marginal disutility of labor, σ the coefficient of relative risk aversion and S_t^C, S_t^L the shocks on intertemporal consumption and labor supply, then the maximization problem of Ricardian households is

$$\max E_t \sum_{t=0}^{\infty} \beta^t S_t^C \left[\frac{C_{R,t}^{1-\sigma}}{1-\sigma} - S_t^L \frac{L_t^{1+\psi}}{1+\psi} \right] \tag{2}$$

subject to the budget constraint

$$P_t\left(1+\frac{\tau_c}{\phi_c^l}\right)(C_{R,t}+I_t)+\frac{B_{t+1}}{R_t^B}=W_tL_t\left(1-\frac{\tau_l}{\phi_t^l}-\tau_p\right)+R_tK_t\left(1-\frac{\tau_k}{\phi_t^k}\right)+B_t. \tag{3}$$

Households make their savings in form of investment into capital goods I_t or in form of bond's purchase issued by government B_{t+1} with the rate of return R_t^B. Every household has to pay taxes on consumption τ_c, on labor income τ_l, on capital income τ_k and also has to pay social security contributions τ_p. ϕ_t^c, ϕ_t^l and ϕ_t^k are the stochastic components of the taxes. Households receive wages W_t for labor, returns from capital R_t and payments from the bonds B_t that matured after one period.

The capital in economy is characterized by law of motion with the rate of depreciation δ,

$$K_{t+1}=(1-\delta)K_t+I_t. \tag{4}$$

Shocks presented to the model have the rules of movement

$$\log S_t^i=(1-\rho_{si})\log S_{ss}^i+\rho_{si}S_{t-1}^i+\epsilon_{si,t},\quad i\in\{C,L\} \tag{5}$$

$$\log\phi_t^j=(1-\rho_j)\log\phi_{ss}^j+\rho_j\phi_{t-1}^j+\epsilon_{j,t},\qquad j\in\{c,l,k\} \tag{6}$$

where $\epsilon_{si,t}$, $\epsilon_{i,t}$ are exogenous shocks and ρ_{si}, ρ_i are autoregressive components.

Inactive workers are the retired ones and are not allowed to make savings. They do not maximize their intertemporal utility function, they are just limited by the value PEN of the benefits received from government and therefore it applies

$$PEN=\left(1+\frac{\tau_c}{\phi_t^c}\right)P_tC_{NR,t}. \tag{7}$$

Firms

The final good of economy is created in two stages. At the beginning there is a *wholesale sector* representing number of firms in monopoly competition that produce different intermediate goods. Firstly, the wholesalers choose the quantities of production factors taking the prices as given in order to minimize their costs

$$\min_{L_{j,t},K_{j,t}} W_tL_{j,t}+R_tK_{j,t} \tag{8}$$

subject to the technology

$$Y_{j,t}=A_tK_{j,t}^{\alpha}L_{j,t}^{1-\alpha}, \tag{9}$$

with the share of capital α and the productivity A_t characterized by the law of motion

$$\log A_t=(1-\rho_A)\log A_{ss}+\rho_A\log A_{t-1}+\epsilon_{A,t} \tag{10}$$

where $\epsilon_{A,t}$ is an exogenous shock and ρ_A is an autoregressive component.

Subsequently they set the optimal price of intermediate good taking the level of output given in order to maximize their profits

$$\max_{P_{j,t}} P_{j,t}Y_{j,t}-W_tL_{j,t}-R_tK_{j,t}. \tag{11}$$

The second stage of production is a *retail industry*, a single firm facing perfect competition that produces a single good for consumption by aggregating the intermediate goods. The firm maximizes its profit taking the prices of goods from wholesalers as given

$$\max Y_{j,t}=P_tY_t-\int_0^1 P_{j,t}Y_{j,t}\,\mathrm{d}j \tag{12}$$

using the following technology

$$Y_t=\left(\int_0^1 Y_{j,t}^{\frac{\varphi-1}{\varphi}}\mathrm{d}j\right)^{\frac{\varphi}{\varphi-1}}, \tag{13}$$

where $Y_{j,t}$ is the intermediate product and φ is the elasticity of substitution between intermediate goods.

The prices in economy are subject to change. In line with Calvo [2] only a randomly selected fraction $(1-\theta)$ of the wholesalers is able to set the optimal price $P_{j,t}^*$ in every period. The remaining part of the firms θ can use only the price from the previous period $P_{j,t-1}$. The overall price level in economy is equal to

$$P_t=\left[\theta P_{t-1}^{1-\varphi}+(1-\theta)P_t^{*1-\varphi}\right]^{\frac{1}{1-\varphi}}. \tag{14}$$

Government

The government in model acts as fiscal and monetary authority and also takes care of the social security system. Only simple social system with no capitalization is considered, so the pension balance BAL_t is equal to the difference of revenues and expenses

$$BAL_t = \tau_p W_t L_t - PEN \tag{15}$$

Fiscal authority needs to finance the purchases of goods and services. Resources are accumulated from the taxes and issued bonds, which allows the government to finance its spending on debt. The public debt is given as

$$\frac{B_{t+1}}{R_t^B} - B_t = P_t G_t - BAL_t - TAX_t, \tag{16}$$

where the tax revenue TAX_t is equal to

$$TAX_t = \frac{\tau_c}{\phi_t^c} P_t(C_t + I_t) + \frac{\tau_l}{\phi_t^l} W_t L_t + \frac{\tau_k}{\phi_l^k} R_t K_t. \tag{17}$$

The government spending is sensitive relatively to the size of public debt with the sensitivity coefficient χ and it holds

$$G_t - G_{ss} = \chi(B_t - B_{ss}), \tag{18}$$

where G_{ss}, B_{ss} are steady-state levels.

Monetary authority controls the interest rate using simple Taylor rule (Taylor [15]) with the goal of output growth and price stability

$$R_t^B = a(Y_t - Y_{ss}) + b(\pi_t - \pi_{ss}) + R_{ss}^B, \tag{19}$$

where a, b are respective sensitivity coefficients, π_t is inflation rate and Y_{ss}, R_{ss}^B, π_{ss} are steady-state levels.

3.1 Calibration

To perform the stochastic analysis of economy's reaction to the shock in consumption taxes it is needed to assign values to structural parameters. The model is calibrated for Australian economy in 2014, which corresponds to the latest available data. Structural parameters are obtained from economic literature, other DSGE works and papers or calculated from available statistics. Following Table 1 presents the values, description and source for parameters reflecting the Australian economy

Parameter		description	Source
σ	1.17	rel. risk aversion coeff.	Gandelman and Hernandez-Murillo[5]
ψ	1	marg. disutility of labor	Sheen and Wang [16]
τ_c	0.14	imp. tax rate on consump.	OECD [11]
τ_k	0.257	imp. tax rate on capital	Carey and Tchilinguirian [3]
τ_l	0.277	imp. tax rate on labor	OECD [13]
τ_p	0.015	tax on payroll	OECD [11]
ω	0.22	pension beneficiaries	OECD [8]
PEN	0.035	benefit payments	OECD [10]
φ	0.3	price stickiness index	Cagliarini, Robinson and Tran [1]

Table 1 Structural parameters

4 Results

In this section we analyse the effects of positive one unit size shock in consumption taxes using impulse response functions generated for 30 periods - see Figure 3 for the most important variables. There is no significant effect on prices of final product and production factors, wages and returns from capital. The increase caused by shock diminishes after two periods and therefore we do not display them.

Changes in consumption

Consumption is aggregated from Ricardian and non-Ricardian households. The shock in consumption taxes decreases the consumption of active workers who can observe the impact directly and adjust their behavior as their budget is lower. Their welfare is therefore decreased.

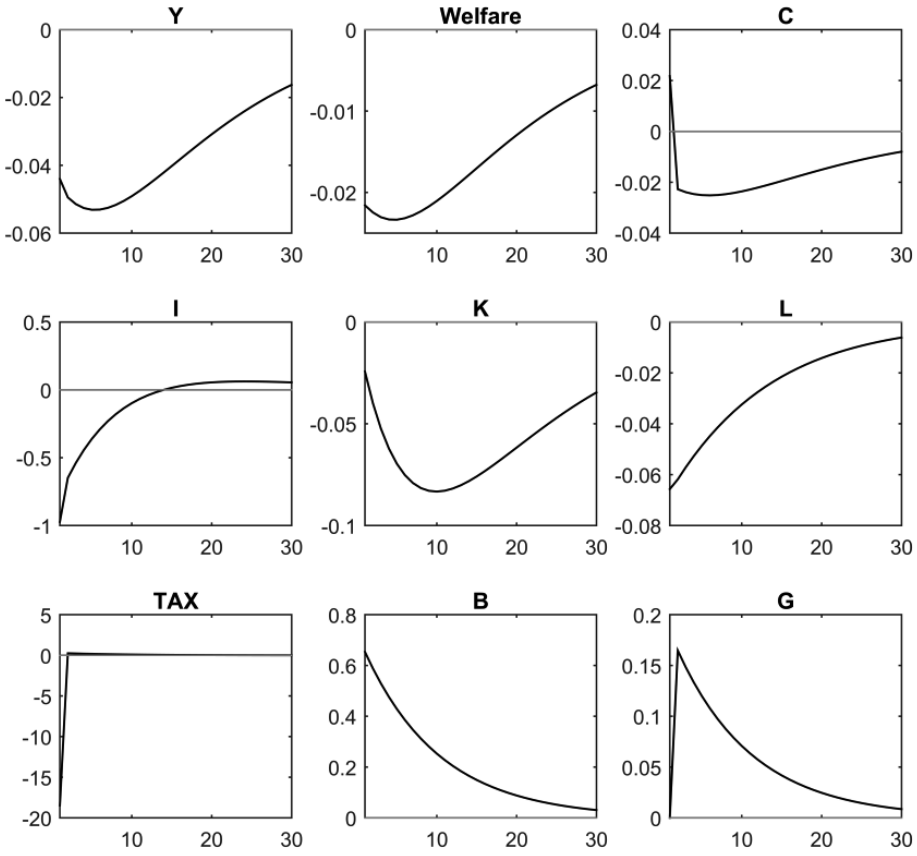


Figure 3 Impulse response functions

Inactive workers increase their consumption in first periods, but this positive change diminishes quickly into downturn and is reflected by the steep line in the first simulated periods. We can explain it with the use of expectations. Inactive workers are limited by payments from government, they can expect that due to increased taxes government could raise the payments and increase their consumption. But the raise is not in place, their budget is lowered by the tax increase and they have to decrease consumption.

Changes in tax revenue, debt and government spending

The expected rise in tax revenue after increasing taxes is overridden by the decrease of consumption and the overall result is flat. Government expects higher tax collection and increases the spending, but together with insufficient tax collection it results in the increase of debt. Also households cannot buy the needed amount of bonds.

Changes in investments

Investments reacts in negative way which can be explained by two factors. Firstly, household's budget is lowered and they cannot make so much savings. Secondly, a displacement of investments from economy is caused by government spending financed on debt. As it is typical for investments, their reaction is the most sensitive amongst the components of GDP.

Overall we can conclude that increasing consumption taxes itself has negative effect on economy. The output is decreased as well as other important variables. It is not good strategy for policy makers just to increase indirect taxes itself. As it was shown in Konôpková [6] decrease of direct taxes can result in positive effect on economy and consolidation of public finances but at the expense of temporary reduced consumption. In case the government wants to increase the consumption taxes then it should be accompanied by the decrease of direct taxes to compensate the negative impact as it is current trend in fiscal policy.

5 Conslusion

In this paper we examined the effects of increasing the taxes on consumption in Australia using DSGE model. We generated the impulse response functions to analyse the reactions of important variables. In general, the economy

reacts negatively on this type of change. We observe the decrease in output, welfare, consumption, investments and on the other side the increase in spending and debt. The public finances cannot be consolidated this way. Therefore we do not recommend government to present this type of reform itself, but combine it with e.g. the decrease in direct taxes.

Acknowledgements

Supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/1076/2016. This support is gratefully acknowledged.

References

[1] Cagliarini, A. & Robinson, T. & Tran, A.:. Reconciling microeconomic and macroeconomic estimates of price stickiness. *Journal of Macroeconomics*, 2011, **33**(1).

[2] Calvo, G.: Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 1983, **12**(3), 383–398.

[3] Carey, D. & Tchilinguirian, H.: Average Effective Tax Rates on Capital, Labour and Consumption. *OECD Economics Department Working Papers*, 2000, No. 258.

[4] Costa Junior, C. J. & Sampaio, A. V.: Tax Reduction Policies of the Productive Sector and Its Impacts on Brazilian Economy. *Dynare Working Papers RePEc:cpm:dynare:036*, 2014.

[5] Gandelman, N. & Hernandez-Murillo, R.: Risk Aversion at the Country Level. *Working Papers 2014-5*, Federal Reserve Bank of St. Louis, 2014.

[6] Konôpková, Z.: Australia: Is it time for tax reform? In: *Proceedings of the Hradec Economic Days 2017* (P. Jedlička, P. Marešová, I. Soukal), University of Hradec Králové, Hradec Králové, 2017. s. 407–413.

[7] Mamatzakis, E. C.: The Dynamic Responses of Growth to Tax Structure for Greece. *Applied Economics Letters*, 2005, **12**(3): 177-180.

[8] OECD: Labour Force Statistics: Summary tables. *OECD Employment and Labour Market Statistics (database)*, 2017. DOI: 10.1787/data-00286-en.

[9] OECD: National Accounts at a Glance. *OECD National Accounts Statistics (database)*, 2017. DOI: 10.1787/data-00369-en.

[10] OECD: *Pension spending (indicator)*, 2017, DOI: 10.1787/a041f4ef-en.

[11] OECD: Revenue Statistics: Comparative tables. *OECD Tax Statistics (database)*, 2017. DOI: 10.1787/data-00262-en.

[12] OECD: Tax Policy Reform and Economic Growth. *OECD Tax Policy Studies*, 2010, 20.

[13] OECD: *Taxing Wages: tax burden on labour income in 2015 and recent trends.* Retrieved from: `http://www.oecd.org/ctp/tax-policy/taxing-wages-tax-burden-trends-latest-year.html`

[14] OECD: *Trade in goods and services (indicator)*, 2017, DOI: 10.1787/0fe445d9-en.

[15] Taylor, J. B.: Discretion versus Policy Rules in Practise. *Carnegie-Rochester Conference Series on Public Policy*, 1993, **39**(1), 195–214.

[16] Sheen, J. & Wang, B, Z. Assessing labor market frictions in a small open economy. *Journal of Macroeconomics*, 2016, **48**(C).

Spa tourism in Slovakia - analysis of defined aspects

Martina Košíková[1], Anna Šenková[2], Eva Litavcová[3], Róbert Bielik[4]

Abstract. The Slovak spas have a long tradition and even in the present the spas preserve a status of improving health of clients and providing a great place for relaxation as well. The aim of the present article is to evaluate significant attributes of service quality for their clients in the three spa companies in Slovak republic - Bardejov Spa, Spa Brusno and Spa Vyšné Ružbachy. In the selected spas we have carried out a research based on a questionnaire survey in which clients evaluated five attributes, among which we included service in accommodation and catering, medical care, technical security, cultural, social and sport-recreation services and services related to spa facilities as a separate tourist destination (for example the number of parking spaces). The article features comparison of the importance of selected attributes in different spa locations and comparison of locations with regard to individual attributes. Based on the outputs, we found out that locations differentiate in the all five attributes. The article features a detailed listing of identified attribute differences within the locations.

Keywords: spa, ANOVA, spa tourism, satisfaction.

JEL Classification: L83
AMS Classification: 62-07, 62G86

1 Introduction

Spas are becoming a global phenomenon, even though in general we do not have many information and evidence describing its scale. Spa services are the fastest growing leisure industry. Thanks to the quality services and proven therapeutic effects, Slovak Republic has a significant position among the world's countries in providing spa services. Even though Slovakia is not a spa power such as Italy, Germany or France, it is one of the countries with high-quality spa services at least at the same level as its neighboring countries Czech Republic, Hungary, Austria and also remote countries like Switzerland, Greece and Spain.

The spa development can be viewed from different perspectives and disciplines, especially hydrology, history, culture, politics, balneology, climatology. The history of Slovak spa is well described in many professional book publications and promotional materials with professional character [1]. At present, Slovak spa companies are mainly providers of health and medical services with proven therapeutic effects and significant improvement of patients' health. In recent years, the competitive environment has increased the importance of other additionally offered services. The health care in Slovakia is at the high level, especially thanks to the tradition, staff professionalism, and legislation that underlines the minimum requirements for personnel and equipment [10]. The aim of employee activities in all spa companies should be to get clients into the spa by providing excellent treatment and top-notch services regarding one's health condition. From the customer point of view, the quality of the provided services is seen from different perspectives such as environment, staff's skills, ability to work on individual needs and flexible handling of different situations. The key element in competition is to enforce quality factors while providing services in the spa environment. Employees are considered to be the most important factor. Based on the results of our spa company culture research we can state that the most important factor is to support teamwork by involving employees to join events, creating rules, delegating work to employees and creating remuneration system [11].

When providing any service, the most important aspects clients evaluate is the way the service is provided, waiting time, safety and reliability of a treatment, hygiene, marketing, information provision, and ongoing quali-

[1] University of Presov in Presov, Faculty of Management, Department of Mathematical Methods And Managerial Informatics, Konštantínova 16, 08001 Prešov, Slovakia, martina.kosikova@smail.unipo.sk
[2] University of Presov in Presov, Faculty of Management, Department of Tourism and Hotel Management, Konštantínova 16, 08001 Prešov, Slovakia, anna.senkova@unipo.sk
[3] University of Presov in Presov, Faculty of Management, Department of Mathematical Methods And Managerial Informatics, Konštantínova 16, 08001 Prešov, Slovakia, eva.litavcova@gmail.com
[4] University of Presov in Presov, Faculty of Management, Department of Tourism and Hotel Management, Konštantínova 16, 08001 Prešov, Slovakia, bielikr001@gmail.com

ty control [4]. Client satisfaction is defined by the overall quality of provided services, as well as a scale to which one's requirements were met.

In the literature we can find many definitions and models of customer satisfaction and its measurement. Januška [5] is an author of a methodology dealing with customer satisfaction in tourism and is a creator of the satisfaction coefficient. Peters and Waterman [8] perceive satisfaction primary as a significant economic factor focused on the success of an organization. The well-known theoretical model focused on finding out customer satisfaction is the SERVQUAL model, which is based on the GAP model [7].

2 Data and methodology

The objects of our research are as follows: spa companies situated in the Slovak Republic: Bardejovské kúpele, a.s. (Bardejov spa) and Kúpele Vyšné Ružbachy (Spa Vyšné Ružbachy) located in the eastern part of Slovakia and Kúpele Brusno, a.s. (Brusno Medical Spa), the spa in the central Slovakia. Several baths have been suited to the research intention according to the chosen criteria (the providing comprehensive spa care services, the indication list of diseases for which the spa treatment is primarily focused, the localization in a unique natural environment). In the pilot survey, we decided to address these three baths, among other things, due to availability.

A portfolio of research samples consists of spa company clients who are long-term visitors of the given spa. One-day and random visitors were not involved in the study.

The feedback we got from the clients was used to collect information on customer satisfaction in the selected companies. The questionnaire was divided into two different parts. The first one deals with socio-demographic characteristics of clients in spas, the second one focuses on the spa destination and visitor satisfaction with the quality of services.

Most of the items related to the socio-demographic profile of spa were closed, although there were questions that could be answered openly. Questions from the second part of the questionnaire were measured by 5-point Likert-type scale, where the respondents were asked about their level of satisfaction with the quality of services (1 = high quality, 5 = poor quality).

The particular questions in the second part of the questionnaire evaluated five attributes, among which we included service provided in the accommodation and catering, medical care, technical amenities of spa, cultural, social opportunities and sport-recreation services and services related to spa facilities as a separate tourist destination (for example the number of parking spaces). We got 150 valid questionnaires (50 per each spa).

The analyses of collected data were focused on evaluating attributes within each specific spa company as well as differences between the spas. The collected data represent the set of categorical data, therefore we decided to use the nonparametric alternative of the Analysis of Variance (ANOVA) and appropriate pair tests, primarily designed to process such structured data. The normality tests of individual attributes (A1-A5 - service in accommodation and catering, medical care, technical amenities of spa, cultural, social opportunities and sport-recreation services and services related to spa facilities as a separate tourist destination) show that none of these attributes is normally distributed, and considering the amount of respondents ($n = 50$ in each spa) we have selected tests that do not require normal distributions.

2.1 Kruskal-Wallis test

Kruskal-Wallis test is a rank-sum test which serves to test the null hypothesis - that k independent random samples come from the identical populations against the alternative hypothesis that the means of these populations are not all equal [2]. Kruskal-Wallis test (or H test) is a nonparametric alternative to ANOVA that does not require normal distributions [12].

From the composite variation range we will get average ranks of all observations in each group $\overline{r_i}$ and also the average of all rank $\bar{r}$. Then the Kruskal–Wallis statistic can be expressed as [9]

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i (\overline{r_i} - \bar{r})^2 \tag{1}$$

or the convenient computational form of H

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{r_i^2}{n_i} - 3(N-1) \tag{2}$$

where

$\overline{r_i}$ the average rank of all observations in the group
r_i the sum of the ranks in the ith sample,
$\bar{r}$ the average of all ranks
n_i the number of observations in the ith sample,
N total number of observations in all samples combined
k the number of samples

If the null hypothesis is true and each sample has at least five observations, the sampling distribution of H can be approximated closely with a chi-square distribution with $k-1$ degrees of freedom. Consequently, we can reject the null hypothesis that $\mu_1 = \mu_2 = \cdots = \mu_k$ and accept the alternative that the μ's are not all equal at the level of significance α, if $H > \chi^2_{1-\alpha}(k-1)$[2]. The chi-square approximation is acceptable when the group sample size $n_i > 5$ with $k \geq 3$ [9].

3 Analysis and results

3.1 Analysis of differences between attributes within spa centers

We used Wilcoxon Matched Pairs Signed-rank Test that compares the differences between the individual attributes within a particular spa. Unlike the t test for paired data, the Wilcoxon signed-ranks test does not require normal distributions [12]. Instead, it tests the hypothesis that the scores for two variables were drawn from the same distribution. The test:

$$Z = \frac{min(S_p, S_n) - (n(n+1)/4)}{\sqrt{n(n+1)(2n+1)/24 - \sum_{j=1}^{l}(t_j^3 - t_j)/48}} \tag{3}$$

where

n number of cases with non-zero differences
l number of ties
t_j number of elements in the j-th tie, $j = 1, \ldots, l$

Table 1 describes average and median values in individual attributes of the spa center. A lower average value (or median) means higher satisfaction with the provided service.

	A1		A2		A3		A4		A5	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Bardejov	1,760	1,333	2,295	2,500	1,848	1,800	1,845	1,750	1,996	2,200
Brusno	1,453	1,333	1,875	1,750	1,616	1,600	1,570	1,500	1,492	1,400
Vyšné Ružbachy	1,550	1,500	2,865	2,625	2,152	2,200	2,685	2,875	2,180	2,400

Table 1 Average and median attribute values of individual baths

Following tables present the results of Wilcoxon Matched Pairs Signed-rank Test. Moreover, they show in which attributes the evaluations differ. Each table (Table 2-4) represents individual spa place (Bardejov, Brusno, Vyšné Ružbachy) with original attributes (left part) as well as modified (ref. Attr.) attributes. Attr. 1-5 were created due to small number of possible values which have de facto ordinal character.

Spa Bardejov											
Attribute (A)	A1	A2	A3	A4	A5	Attribute (Attr.)	Atr1	Atr2	Atr3	Atr4	Atr5
A1	-	***0,000***	0,195	0,127	***0,002***	Attr1	-	0,129	***0,043***	0,536	***0,000***
A2	-	-	***0,000***	***0,000***	***0,035***	Attr2	-	-	***0,001***	***0,001***	***0,000***
A3	-	-	-	0,972	0,573	Attr3	-	-	-	0,238	***0,002***
A4	-	-	-	-	0,226	Attr4	-	-	-	-	***0,000***
A5	-	-	-	-	-	Attr5	-	-	-	-	-

Table 2 Wilcoxon signed-rank test – Spa Bardejov

Spa Brusno											
Attribute (A)	A1	A2	A3	A4	A5	Attribute (Attr.)	Atr1	Atr2	Atr3	Atr4	Atr5
A1	-	***0,000***	***0,005***	0,109	0,161	Attr1	-	0,099	***0,014***	0,186	0,200
A2	-	-	***0,000***	***0,000***	***0,000***	Attr2	-	-	0,484	0,994	***0,013***
A3	-	-	-	0,608	***0,032***	Attr3	-	-	-	0,245	***0,000***
A4	-	-	-	-	0,260	Attr4	-	-	-	-	***0,043***
A5	-	-	-	-	-	Attr5	-	-	-	-	-

Table 3 Wilcoxon signed-rank test – Spa Brusno

Spa Vyšné Ružbachy											
Attribute (A)	A1	A2	A3	A4	A5	Attribute (Attr)	Atr1	Atr2	Atr3	Atr4	Atr5
A1	-	***0,000***	***0,000***	***0,000***	***0,000***	Attr1	-	***0,000***	***0,000***	***0,000***	***0,000***
A2	-	-	***0,000***	***0,028***	***0,000***	Attr2	-	-	0,181	0,847	0,316
A3	-	-	-	***0,000***	0,766	Attr3	-	-	-	***0,026***	***0,015***
A4	-	-	-	-	***0,000***	Attr4	-	-	-	-	0,124
A5	-	-	-	-	-	Attr5	-	-	-	-	-

Table 4 Wilcoxon signed-rank test – Spa Vyšné Ružbachy

The results show which attributes within the individual spas show statistically significant differences. The most different evaluation of original attributes was found in Vyšné Ružbachy spa. There is a statistically significant difference between almost all attributes except A3 and A4.

3.2 Analysis of differences between spas' individual attributes

The analysis of differences between individual spas was carried out separately for each attribute. To do so we used two tests – Kruskal-Wallis test and Median test. In these tests individual attributes were considered as dependent variables. The attributes were outlined for the three different groups (spa Bardejov, Brusno and Vyšné Ružbachy). Then we determined whether or not there are statistically significant differences in the median of the monitored groups.

Based on the obtained results we note that all attributes (for three spas) feature at least one statistically significant difference.

The results of Kruskal – Wallis test and Median test, representing the analysis of the differences in the attributes of the spa – the so called "average rank" in the composite variation range, are shown in Table 5.

	A1	A2	A3	A4	A5
Bardejov	80,37	74,02	72,51	65,86	83,16
Brusno	63,74	50,28	57,85	51,54	45,61
Vyšné Ružbachy	82,39	102,20	96,14	109,10	97,73
Kruskal-Wallis test: H	5,6842	36,5036	20,1360	48,2144	38,8816
Median test: Chi-Square	1,7322	24,0079	13,5000	30,7200	39,5200

Table 5 Kruskal-Wallis test and Median Test

We also used Kruskal-Wallis test and Median test on the modified attributes. The specific consequential values are shown in Table 6.

	Atr1	Atr2	Atr3	Atr4	Atr5
Bardejov	79,58	64,00	72,79	66,31	85,61
Brusno	62,91	64,94	51,01	50,18	37,30
Vyšné Ružbachy	84,01	97,56	102,70	110,01	103,59
Kruskal-Wallis test: H	6,7673	21,7223	36,6235	51,7116	63,8186
Median test: Chi-Square	8,0893	26,1182	23,4801	46,3167	60,6429

Table 6 Kruskal-Wallis test and Median Test – modified attributes

The Kruskal-Wallis test found significant differences in the level of satisfaction in four attributes of reviews - "medical care", "technical amenities", "cultural, social opportunities and sport-recreation services" and "services related to spa facilities as a separate tourist destination" – p-value = 0,000. Only in one attribute, namely "service in accommodation and catering", satisfaction scores differed between spas, but the Kruskal-Wallis test did not report statistically significant differences (p-value $= 0{,}0583$ for Kruskal-Wallis test and p-value $= 0{,}4206$ for Median Test).

With modified attributes, there is a statistically significant difference in the perceived satisfaction in all attributes. This may be due to the fact that some of these modified attributes contain only one variable.

3.3 Factor analysis of the satisfaction in spas

To obtain a more detailed insight into the results we used the factor analysis. As Garson [3] states, the factor analysis uncovers the latent structure of the set of variables. It reduces attribute space from a larger number of variables to a smaller number of factors and as such is a non-dependent procedure. According to [6], the factor analysis is determined primarily for interval variables and works also well with either ordinal or binary data.

The factor analysis generates a table in which the rows feature the observed raw indicator variables and the columns the factors or latent variables which explain as much of the variance in these variables as possible. The cells in this table are factor loadings, and the meaning of the factors must be induced from seeing which variables are most heavily loaded on which factors [3].

The so-called Rule 10 is met as the ratio criterion of the number of subjects to the number of variables is greater than 10: There should be at least 10 cases for each item used ([3; 6]. Therefore, we adjust the original questions (i.e. individual questions in five attributes) only to those questions that contained answers from the whole scale - from high quality to poor quality. From a complex selection, we have a set of 13 items (questions) within the five attributes.

During the choice of rotation and by comparing different rotated solutions it was confirmed that the most appropriate rotation was the orthogonal Varimax rotation minimizing the number of variables which have high factor loadings in more factors [6].

The factor loadings (the correlation coefficient between a factor and a variable) of the rotated solution are obtained as a result and shown in Table 7. The rotated solution enables us to interpret extracted factors [6].

	Variable	Factor 1	Factor 2	Factor 3	Factor 4
accommodation and catering	Room size and layout of furniture	0,06902	0,07742	-0,11533	***0,83170***
	Professional staff at hotel in spa	0,14408	***0,75299***	-0,08582	0,43823
	Variety of meal offer	0,22409	***0,88107***	0,05462	0,08185
	Offering dietary meals	0,11096	0,43937	0,17999	***0,67895***
medical care	Recommendation of mineral springs by a doctor	0,38323	0,06668	***0,80997***	-0,11570
technical amenities	Outdoor appearance of baths	-0,14106	0,37158	***0,57446***	***0,59902***
	Medical equipment for examinations	0,52583	0,38791	***0,57811***	-0,04538
	Infrastructure in the spa area	0,35599	0,03711	***0,67861***	0,30264
cultural and social events, sport-recreation services	Offer of social - entertainment events	***0,79044***	0,26162	0,40627	-0,00718
	Offer of business services	0,03891	***0,69016***	0,44199	0,12202
	Spiritual opportunities in the spa	0,35968	-0,03732	***0,70802***	0,04706
services related to spa facilities	Number of parking places for one-day tourists, visitors or spa guests	***0,91954***	0,06984	0,14114	-0,04802
	Public lighting of the spa	***0,77035***	0,11584	0,20264	0,37622

Table 7 Factor analysis of the satisfaction in spas

Factor 1 is saturated with items *offer of social - entertainment events, number of parking places* and *public lighting of the spa.* The appropriate term for the first factor is "Accessibility and level of cultural and social services."

Factor 2 is saturated with the items *professional staff, meal offer* and *offer of business services* and the adequate name for it is "Availability and level of catering and business services".

Factor 3 is saturated with the items *recommendation of mineral springs by a doctor, medical equipment for examinations, infrastructure in the spa area* and *spiritual opportunities in the spa* which can be expressed by "The quality and level of spa and therapeutic services."

Factor 4 is saturated with items *room size*, o*ffering dietary meals* and *outdoor appearance of bath.* The appropriate term for the first factor is "The quality and level of technical amenities".

4 Conclusion

Tourism's product is basically customer service. Spa tourism is a specific form of tourism, which means the provision of services on a wider scale, like accommodation and catering, medical care, socio-cultural services, technical amenities of a spa and other services related to spa facilities as separate tourist destination in Slovakia.

This article shows the importance and significance of providing quality services for customers. Doing so can ensure the competitiveness of a company over such companies whose level of provided services is much lower.

This study specifically aimed to evaluate significant attributes of service quality in the spa companies in Slovakia. It shows that there is no statistically significant difference between particular spa companies when it comes to evaluated services like accommodation and catering. However, some other attributes show significant differences. The best-rated spa company was spa Brusno in (all five attributes). Surprisingly enough, the primary purpose of a spa (the regeneration of the body and the related quality of medical care) proved to be the worst rated attribute in all selected spas.

Creating a new, modern and first of all competitive spa in the conditions of our country is a challenge for the next generations (provided Slovakia will still have natural healing resources). It is up to us to take advantage of our possibilities and make sure that we provide clients with quality services in all the above attributes.

Acknowledgements

Supported by the grant VEGA No. 1/0412/17 of the Scientific Grant Agency of Ministry of Education SR and Slovak Academy of Sciences.

References

[1] Eliášová, D.: *Slovenské kúpeľníctvo v 20. storočí*. Vydavateľstvo Ekonóm, Bratislava, 2009.
[2] Freund, J. E., Williams, F. J., and Perles, B. M.: *Elementary bussiness statistics: the modern approach.* Prentice Hall, 1988.
[3] Garson, D.: Factor Analysis: Statnotes. Retrieved March 22, 2008, from North Carolina State University Public Administration Program, http://www2.chass.ncsu.edu/garson/pa765/factor.htm.
[4] Gúčik, M., Gajdošík, T., and Lencsésová, Z.: *Kvalita a spokojnosť zákazníka v cestovnom ruchu*. Wolters Kluwer, Bratislava, 2016.
[5] Januška, Ľ.: Spokojnosť subjektu s pobytom v stredisku cestovného ruchu. In: *Ekonomická revue cestovného ruchu,* 1981, Roč. 14, č. 4. ISSN 0139-8660.
[6] Litavcová, E., Bucki, R., Štefko, R., Suchánek, P., and Jenčová, S.: Consumer′s Behavior in East Slovakia after Euro Introduction during the Crisis. *Prague Economic Papers,* Vol. 24, no. 03/2015, 332–353.
[7] Mateides, A., and Ďaďo, J.: *Služby.* Vydavateľstvo Epos, Bratislava, 2002.
[8] Peters, T. J., Waterman, R. H.: In search of Excellence. Harper and Row, New York.
[9] Ramachandran, K. M., and Tsokos, C. P.: *Mathematical Statistics with Applications in R.* Academic Press. 2015.
[10] Šenková A., and Matušíková, D.: Súčasný stav ponuky kúpeľníckych zariadení na Slovensku. In: *Aktuální trendy v lázeňství, hotelnictví a turismu* [elektonický zdroj]: sborník příspěvků: mezinárodní vědecká conference. Opava: Slezská univerzita Opava, 2012.
[11] Šenková, A., Šambronská, K., Mitríková, J., and Maťková, S.: Corporate culture as a tool for increasing employee motivation. In: *Polish journal of management studies.* 2016, Vol. 13, no. 2, online, pp. 131-141,
[12] Triola, M. F.: *Elementary statistics.* Addison-Wesley Publishing Company, Inc., 1992.

Optimal choices among ethic assets of the Italian market

Noureddine Kouaissah[1], Sergio Ortobelli[2], Marianna Cavenago[3]

Abstract. In this paper, we explore and examine the performance of Italian socially responsible investments (SRI). In particular, we use different portfolio strategies to evaluate the financial performance of the Italian SRI. In this context, we propose a new portfolio strategy that maximizes the Sharpe ratio under the third-order stochastic dominance constraints. Applying the portfolio optimization methods to the Italian ethical funds/indexes, we are able to answer several questions that arise with SRI opportunities. Initial empirical results from Italian market are in contrast with the corporate social responsibility (CSR) theory conclusion, which confirms that SRI are performing much better than their conventional counterparts. This may well be linked to the characteristics and the peculiarity of the Italian market, e.g. recentness and dimensionality. The proposed empirical analysis allows us to understand the complexity and dynamics of the Italian SRI and to evaluate their performance.

Keywords: financial performance, SRI, portfolio optimization, CSR theory.

JEL Classification: G11, C44, G10.
AMS Classification: 28A25.

1 Introduction

Socially responsible investments (SRI) have been significantly increased over the years as measured by a number of the asset under management. Socially responsible investors consider both financial and social criteria in their investment decisions in order to ensure that securities are selected according to their values and beliefs. The use of socially responsible screens is not a new concept, see Sauer (1997). In the early 1900's, socially responsible investors consciously avoided companies that were involved in the production of weapons, alcohol, tobacco, and gambling. More recently, social concerns have expanded to include different principles, such as Environmental, Social and Governance (ESG) considerations and human rights, in investment decisions to achieve both positive societal impact and competitive financial returns.

While many investors have clearly chosen their portfolios so that maximize their expected utility, many others are probably more than willing to give up some return to invest according to their values and beliefs. In this context, Guerard (1997) attempts to examine whether there exists a cost to being socially responsible investing or not. Furthermore, Diltz (1995)) and Sauer (1997) examine whether social screening affects the portfolio performance. Conceptually, the debate on the performance of SRI has been centered on the tradeoff between the arguments of diversification and the benefits of CSR. On the one hand, several authors, in accordance with the postulates of modern portfolio theory (see Markowitz 1952), argue that screening criteria limit the possibility of creating a well-diversified portfolio. On the other hand, opponent scholars confirm that ethical companies obtain better performance than traditional companies in the long term as a result of their positive stakeholder relationship. This would allow SRI outperform conventional investments in a later phase of development in line with stakeholder theory, for more details see Freeman et al. (2010).

Bearing this in mind, we use two well-known strategies (i.e. uniform portfolio and Sharpe ratio) and an innovative strategy called Sharpedom to examine the Italian SRI. More specifically, we propose an optimization method based on the Sharpe ratio under the third stochastic dominance (TSD) constraints. We use the TSD because it is rather easy to verify and accepted by financial economists based on compelling theoretical and empirical parameters, see among others Post and Kopa (2016). TSD is a less restrictive form than second-degree stochastic dominance (SSD) since it considers preference ordering only for those risk-averse investors who exhibit decreasing risk aversion (more likely to present socially responsible investors' preferences). Then, we compare the performance of the proposed strategy to that of the classical Sharpe ratio and the uniform portfolio. Finally, through an empirical analysis, we examine the performance of the Italian ethical funds/indexes.

[1] University of Bergamo, Department of Management, Economics and Quantitative Methods, via dei Caniana 2 Bergamo, Italy, noureddine.kouaissah@unibg.it .
[2] University of Bergamo, Italy & VSB-TU of Ostrava, Czech Republic, sergio.ortobelli@unibg.it .
[3] University of Bergamo, Department of Management, Economics and Quantitative Methods, via dei Caniana 2 Bergamo, Italy, mariannacavenago@gmail.com .

The rest of the paper is organized as follows. Section 2 presents the suggested portfolio selection methodologies. In section 3, we discuss the dataset used and the ex-post empirical results. Finally, section 4 concludes.

2 Portfolio selection problems

The portfolio selection problem addresses how investors allocate their wealth across the available securities in order to realize a profitable level of returns. Clearly, in the literature, we can find many different performance measures, see among others Rachev et al. (2008). For example, Rachev ratio is the ratio between the average of the largest earnings and the average of the largest losses. Let us briefly formalize the portfolio performance measures (e.g. uniform portfolio and Sharpe ratio) that are used in the empirical analysis section.

Uniform portfolio, see DeMiguel et al. (2009) and Pflug et al. (2012), defined as one in which a fraction 1/*n* of wealth is allocated to each of the *n* assets available for investment at each recalibration time. This strategy, also known as naïve portfolio diversification, is easy to implement because does not rely either on optimization techniques or on the estimation of the moments of asset returns. Several comparative studies show that uniform portfolio is hard to beat as an investment strategy. Therefore, in the empirical analysis, we merely use it as a benchmark to assess the performance of the other portfolio rules.

Sharpe ratio (1994). The Sharpe ratio is used to characterize how well the return of an asset compensates the investor for the risk taken. The Sharpe ratio computes the price for the unity of risk and calculated by subtracting the risk-free rate from the rate of return of the portfolio and dividing the result by the standard deviation of the portfolio returns. Formally:

$$SR(x'z) = \frac{E(x'z) - z_b}{\sigma_{x'z}}, \tag{1}$$

where, $E(x'z)$ is the portfolio expected returns, z_b is the risk-free return, $x'z = \sum_{i=1}^{n} x_i z_i$ denotes the returns of a portfolio, $x = [x_1, \dots, x_n]'$ is the vector of the positions taken in the n assets and $\sigma_{x'z}$ is the portfolio standard deviation.

Generally maximizing the Sharpe ratio, we get a market portfolio that should be optimal for non-satiable risk-averse investors (see, for instance, Rachev et al. 2008). As an innovative strategy, we propose an optimization method which maximizes the Sharpe ratio under the third-order stochastic dominance (TSD) constraints. In particular, as constraints, we consider the optimal portfolios that dominate a given benchmark (e.g. uniform portfolio) in terms of TSD. For simplicity, we refer to this strategy as Sharpedom. The TSD is rather easy to verify and accepted by financial economists based on compelling theoretical and empirical arguments, see among others Post and Kopa (2016).

The economic meaning of the TSD can be explained as follows. Denote $\mathcal{U}_3$ the set of all utility functions that are non-decreasing, concave and have a non-negative third derivative. Thus, $\mathcal{U}_3$ represents the class of non-satiable, risk-averse investors (with decreasing risk aversion) who prefer positive to negative skewness. We say that a portfolio X dominates the benchmark Y in the sense of TSD, $X \succeq_{TSD} Y$, if any investor with a utility function belongs to $\mathcal{U}_3$ does not prefer Y to X. In other words, $X \succeq_{TSD} Y$, if $Eu(X) \geq Eu(Y), \forall\, u \in \mathcal{U}_3$; for deeper discussion see, among others, Rachev et al. (2008) and Whitmore (1970).

Therefore, when no short sales are allowed ($x_i \geq 0$) and, to guarantee a diversification, we cannot invest more than a fixed percentage θ in any asset ($x_i \leq \theta$), we assume that investors will choose the market portfolio solution of the following constrained optimization problem:

$$\begin{aligned} &max_x \ \rho(x'z) \\ &\text{s.t.} \quad \textstyle\sum_{i=1}^{n} x_i = 1 \\ &x_i \geq 0;\ x_i \leq \theta;\ i = 1, \dots, n \end{aligned} \tag{2}$$

where $\rho(x'z)$ is the Sharpe ratio (1). While adding the TSD constraint to the optimization problem (2) we obtain the Sharpedom strategy as follows:

$$\begin{aligned} &max_x \ \rho(x'z) \\ &\text{s.t.} \quad \textstyle\sum_{i=1}^{n} x_i = 1 \\ &x'z \succeq_{TSD} \text{"}1/n\text{"} \\ &x_i \geq 0;\ x_i \leq \theta;\ i = 1, \dots, n \end{aligned} \tag{3}$$

where $\rho(x'z)$ is the Sharpe ratio (1) and $x'z \succeq_{TSD}$ "$1/n$" represents the TSD constraint with respect to the uniform portfolio.

3 Data and Ex-post empirical analysis

In this section, we propose an empirical analysis of several admissible portfolio optimization problems using Italian ethical investments (funds/indexes). Firstly, the dataset description is provided (Sect. 3.1). Secondly, we evaluate three main portfolio strategies on the same dataset (Sect. 3.2). Finally, we discuss the results and summarize our conclusion on the performance of SRI.

3.1 Dataset description

In order to address the crucial question whether ethical investing worth in the Italian market, we examine the Standard Ethics Italian index, composed of the 40 largest Italian listed companies with a Standard Ethics Rating; Standard Ethics Italian Banks Index that includes all Italian banks listed on the Italian Stock Exchange (currently 22), and then the Standard Ethics Mid Italian Index, consisting of medium size listed companies with a Standard Ethics Rating. In particular, the Mid Italian Index includes 12 companies that have a market capitalization greater than 250 million euros. For each index, we first recuperate the component list from the Standard ethics website. Then, we collect the dataset contained in Thomson Reuters DataStream for the period January 1, 2000, through July 11, 2015. Finally, in order to examine the ethical mutual funds' performance, we consider Etica Sgr which is an Italian asset management company founded in 2000 and is developing, promoting and managing exclusively SRI. First, we recuperate the component list. Then, we obtain the dataset contained in Thomson Reuters DataStream for the period January 21, 2007, through July 11, 2015.

3.2 Ex-post empirical analysis

In this section, we report the main results obtained optimizing three different performance measures (i.e. Sharpe ratio, uniform portfolio and Sharpedom strategy) on the different ethical dataset as summarized in the previous section. We use a moving average window of 125 trading days for the computation of each optimal portfolio and we recalibrate the portfolio every 20 trading days. We assume that short sales are not allowed and, in order to guarantee enough diversification, no single stock may comprise more than 20 % of the portfolio. We also consider proportional transaction cost of 20 basis points. Starting with an initial wealth $W_0 = 1$ that we invest at the beginning of each period, we evaluate the ex-post wealth sample paths obtained from the considered strategies (Sharpe ratio and Sharpedom). Thus, at the *k*-th recalibration time, the following steps are performed for Sharpe and Sharpedom strategies:

Step 1. Determine the market portfolio $x_M^{(k)}$ that maximizes the performance measure, i.e. the solution of the following optimization problem:

$$\begin{aligned} &max_x \ \rho(x'z) \\ &\text{s.t. } \textstyle\sum_{i=1}^{n} x_i^{(k)} = 1, \\ &0 \le x_i^{(k)} \le 0.2 \quad i = 1, \dots, n \end{aligned} \tag{4}$$

Here the performance measure $\rho(x'r)$ is the Sharpe ratio (1). While including the constraint $x'z \succeq_{TSD}$ "$1/n$" we obtain the Sharpedom strategy as (3).

Step 2. Since we recalibrate the portfolio every 20 days, we calculate the ex-post final wealth as follows:

$$W_{t_{k+1}} = \left(W_{t_k} - t.c._{t_k}\right)\left(x_M^{(k)}\right)' z_{(t_{k+1})}^{(expost)}, \tag{5}$$

where $t.c._{t_k}$ are the proportional transaction costs of 20 basis points and $z_{(t_{k+1})}^{(expost)}$ is the vector of observed gross returns in the period between t_k and t_{k+1}, such that $t_{k+1} = t_k + 20$.

Steps 1 and 2 are repeated for both performance measures until the observations are available. The results of these analyses and those of the uniform portfolio on Standard Ethics are reported in Table 1, which contains 3 panels according to which dataset is used. In particular, each panel reports summary statistics (mean, standard deviation, skewness, kurtosis, VaR 5%, CVaR 5%, final wealth) of the ex-post returns of the three strategies. In addition, we compute the Sharpe ratio and the performance measure $STARR_\alpha$, defined by Martin et al. (2003),

with a confidence level $1 - \alpha = 95\%$. STARR allows us to overcome the drawbacks of the standard deviation as a risk measure (Artzner et al. (1999)) and focuses on the downside risk. [4]

Table 1: Average of some statistics of the ex-post returns obtained by maximizing the Sharpe ratio, Sharpedom and uniform strategy using Standard Ethics components.

Panel A: Standard Ethics Italian index

	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5 %	Final W	Sharpe	STARR
Sharpe	0.018%	1.561%	-0.0530	3.7430	2.526%	3.907%	1.2316	1.123%	0.449%
SRdom	0.019%	1.350%	-0.0440	3.3805	2.205%	3.668%	1.4857	1.425%	0.524%
Uniform	0.031%	1.511%	0.0579	3.1419	2.466%	4.275%	2.2034	2.084%	0.737%

Panel B: Standard Ethics Italian Banks Index

	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5 %	Final W	Sharpe	STARR
Sharpe	0.011%	1.429%	0.0985	8.0260	2.199%	5.108%	1.0229	0.758%	0.212%
SRdom	0.001%	1.203%	-0.0266	10.499	1.860%	4.242%	0.7475	0.005%	0.002%
Uniform	-0.002%	1.465%	0.0536	3.6704	2.427%	5.339%	0.6007	-0.148%	-0.041%

Panel C: Standard Ethics Mid Italian Index

	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5 %	Final W	Sharpe	STARR
Sharpe	0.035%	1.641%	0.0978	7.5762	2.567%	3.946%	2.2946	2.109%	0.877%
SRdom	0.039%	1.325%	0.0744	7.2180	2.146%	3.474%	3.2948	2.953%	1.127%
Uniform	0.036%	1.132%	-0.3078	4.2153	2.023%	3.083%	3.0116	2.894%	1.156%

This Table contains eight different statistics on the ex-post log returns obtained with three different portfolio strategies (i.e. Sharpe ratio, Sharpedom strategy and uniform portfolio).

Panel A) reports the results for Standard Ethics Italian index. From this panel, we clearly observe that the uniform portfolio achieves the greatest average, final wealth, Sharpe ratio (mean/St. dev.) and STARR performance. Interestingly, Sharpedom strategy presents the lowest risk (standard deviation, VaR 5%, CVaR 5%). In particular, Sharpedom strategy dominates the uniform portfolio during the period of high losses. This is an important result compared to the uniform portfolio, which is a strategy with lower losses by definition. Moreover, if we consider transaction costs per operation (investment/disinvestment), which should be higher for uniform portfolio, we could have more comparable results between both strategies.

Panel B) displays the results based on Standard Ethics Italian Banks Index dataset. Generally, this index is more pertinent to the CSR principles (including all Italian banks listed on the Italian Stock Exchange). From this panel, we observe that both Sharpedom strategy and Sharpe ratio are performing much better than the uniform portfolio in terms of the mean, final wealth, Sharpe ratio (mean/St. dev.) and STARR performance. However, due to the recent financial turmoils, principally subprime crisis and then European debt crisis, all strategies lose significant values. As was expected, these results reflect the sensitivity of the banking sector to the sovereign debt crisis.

Panel C) presents the results using the Standard Ethics Mid Italian Index. This index represents 12 companies that have a market capitalization greater than 250 million euros. According to this panel, the Sharpedom strategy presents the greatest average, final wealth and Sharpe ratio (mean/St. dev.). While the uniform portfolio shows the highest risk (standard deviation, VaR 5%, CVaR 5%). On the other hand, Sharpe ratio achieves the worst results among the three strategies.

To enrich our analysis, we consider Etica Sgr which is an Italian asset management company founded in 2000 and is developing, promoting and managing exclusively SRI. Figure 1 reports the sample paths of the ex-post wealth obtained with the three examined portfolio strategies (i.e. Sharpe ratio, Sharpedom strategy and uniform portfolio) on Etica Sgr fund components.

[4] STARR ratio is not a symmetric measure of risk when returns present heavy-tailed distributions, see Martin et al. (2003).

Figure 1: Ex-post wealth obtained with Sharpe ratio, Sharpedom strategy and uniform portfolio using Etica Sgr fund dataset.

From Figure 1, we surprisingly observe that Etica Sgr benchmark is performing much better than the three strategies. Furthermore, both strategies uniform portfolio and Sharpe ratio are performing much better than the proposed strategy Sharpedom, which dominates them only during the period of subprime crisis (2008). Towards completing this analysis, we compare the performance of the examined benchmarks over the same period (even we are aware of the dimensionally and diversification effect between different benchmarks). We aim to answer the question whether or not Etica Sgr benchmark outperforms the Raiffeisen Azionario Sostenibile, MSCI Social and S&P 500 index.

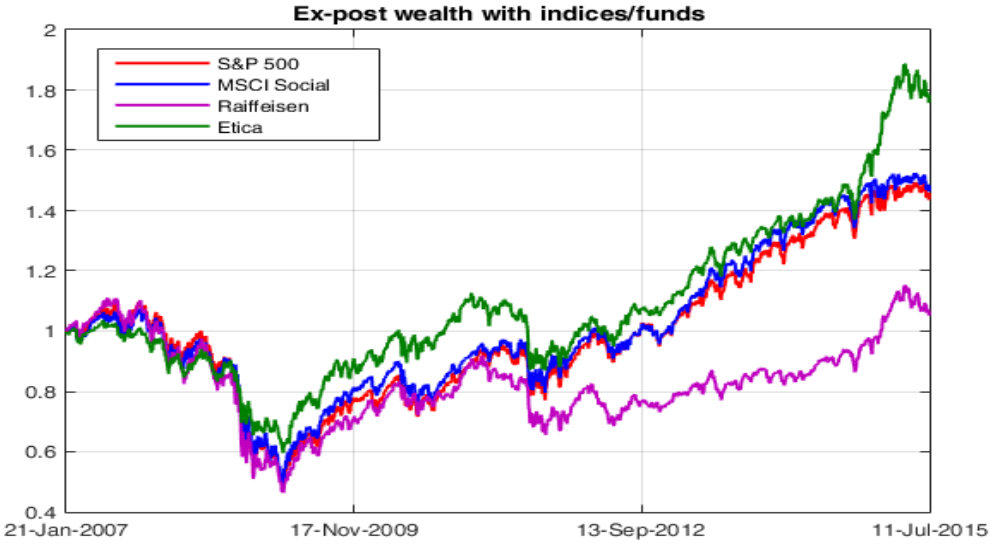


Figure 2: Ex-post wealth obtained investing on four different indexes/funds (i.e. S&P 500, MSCI Social, Raiffeisen Azionario Sostenibile fund and Etica Sgr fund).

From Figure 2, we generally observe that Etica Sgr mutual fund dominates all other benchmarks, especially after the subprime crisis, while the Raiffeisen Azionario Sostenibile mutual fund is the worst among all. However, before the subprime crisis, we can see the opposite results, which means that Etica Sgr was slightly underperformed the other benchmarks. Whereas American indices allocate between the previous ones.

Summarizing, SRI have grown rapidly around the world over the last decade, reflecting the increasing awareness of investors to ESG issues. This paper provides an overview of the SRI performance in the Italian market. In particular, using sound portfolio strategies and some statistics, these studies allow us to explicitly understand the complexity and dynamics of the SRI. We find that dimensionality and recentness have some effect on the Italian market.

4 Conclusion

In this paper, we discuss and empirically analyze the financial performance of the SRI from the portfolio theory perspective. Optimal portfolios are derived using three different strategies (i.e. Sharpe ratio, Sharpedom, and uniform portfolio). In particular, we propose Sharpedom as a new portfolio strategy that optimizes the Sharpe ratio under the TSD constraints. The objective is to evaluate and examine the financial performance of the Italian SRI.

We generally observe that the ex-post results from Italian market are in contrast with the corporate social responsibility (CSR) theory conclusion, which confirms that SRI are performing much better than the traditional investments. This may well be linked to the characteristics and the peculiarity of the Italian market, e.g. recentness and dimensionality of Italian funds/indices. The proposed empirical analysis allows us to better understand the dynamics of the ethical investments and to evaluate their performance. Overall, the results reveal the complexity SRI investors face in finding ethical investments that meet their non-financial goals.

Further research could involve theoretical and empirical studies. On the one hand, other performance measures could be explored in order to better address investor's preferences. It would be interesting to examine the SRI performance, according to different risk aversion (i.e. nonsatiable risk-averse, nonsatiable risk-seeking, neither risk-averse nor risk-seeking). On the other hand, it could be possible to apply the proposed studies to other markets, considering their specific structure.

Acknowledgements

The paper has been supported by the Italian funds MURST 2016. The research was also supported through the Czech Science Foundation (GACR) under project 17-19981S, by VSB-TU Ostrava under the SGS project SP2017/32.

References

[1] Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203-228.

[2] Davidson, R., Jean-Yves, D. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, 68(6), 1435-1464.

[3] DeMiguel, V., Garlappi, L., Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22, 1915-1953.

[4] Diltz, J. D., (1995). Does social screening affect portfolio performance? *Journal of Investing*, 4(1), 64-69.

[5] Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., de Colle, S. (2010). Stakeholder theory: the state of the art. *Cambridge University Press*, New York.

[6] Guerard, J. B. (1997). Additional evidence on the cost of being socially responsible in investing. *Journal of Investing*, 6(4), 31-36.

[7] Markowitz, H. M. (1952). The utility of wealth. *Journal of Political Economy*, 60, 151-158.

[8] Martin D., Rachev S., Siboulet F., (2003). Phi-alpha optimal portfolios and Extreme Risk Management. *Wilmott Magazine of Finance*, November, 70-83.

[9] Moskowitz, M. (1997). Social investing: the moral foundation. *Journal of Investing*, 6(4), 9-11.

[10] Müller, A., Stoyan, D. (2002). Comparison methods for stochastic models and risks. *John Wiley & Sons*, New York.

[11] Pflug, G., Pichler, A., Wozabal, D. (2012). The 1/N investment strategy is optimal under high model ambiguity. *Journal of Banking and Finance*, 36(2), 410-417.

[12] Post, T., and Kopa, M. (2016). Portfolio Choice based on Third-degree Stochastic Dominance. Forthcoming in *Management Science*. http://dx.doi.org/10.1287/mnsc.2016.2506.

[13] Rachev, S., Stoyanov, S., Fabozzi, F., J. (2008).Advanced stochastic models, risk assessment, and portfolio optimization. *Wiley & Sons, Inc*., Hoboken, New Jersey.

[14] Sauer, D. A. (1997). The impact of social-responsibility screens on investment performance: evidence from the Domini 400 Social Index and Domini Equity Mutual Fund. *Review of Financial Economics*, 6, 137-49.

[15] Sharpe, W.F. (1994). The Sharpe ratio. *Journal of Portfolio Management*. Fall, 45-58.

[16] Whitmore, G., A. (1970). Third-degree Stochastic Dominance. *American Economic Review*, 60, 457-59.

Transformation of task to locate a minimal Hamiltonian circuit into the problem of finding the Eulerian path

Petr Kozel[1], Václav Friedrich[2], Šárka Michalcová[3]

Abstract. If we want to solve tasks aimed at designing optimal route of utility vehicle, the attention can be divided into two sets of tasks. Firstly, it is a task aimed at vertex service of transport network, when primary problem is to locate a minimum of Hamiltonian circuit. Secondly, we focus on services of edges in a transport network, where the default problem is to find the Eulerian path. In this paper, we shall present a possibility of transformation of the task to find minimum Hamiltonian circuit in order to locate the Eulerian path, which is modified for a wider transport network. In addition, some complementary network for an edge service can be used in this case. This network allows an effective transits of service vehicle to reduce unproductive distance traveled. For the sake of accuracy, a mathematical programming approach will be used for the above mentioned problem. Furthermore, real data experiments will be presented, too.

Keywords: mathematical programming, minimum Hamiltonian circuit, Eulerian path, minimum spanning tree, mathematical models.

JEL classification: C61
AMS classification: 90C05

1 Introduction

Basically, tasks aimed at designing optimal vehicle routes can be solved by two principal approaches. Firstly, we search for the optimal route or the optimal cycle with the aid of vertex operation in a graph. With regard to a practice, the frequent issue of collection distribution tasks are customers requirements located at the vertices of the transport network, for instance the so-called Traveling Salesman Problem. Secondly, optimal route can be determined by edge servicing, which in practice is mostly about road maintenance and services. Generally, these trails must be either served or they should encompass some level of transport intended to be distributed. For example, the latter approach is theoretically captured in the Chinese Postman Problem.

In general, the common optimization criterion of the above mentioned approaches is total distance traveled by the vehicle, mostly denominated in kilometers. Overall, both tasks can be solved by three basic types of transport networks, graphs respectively: undirected transport network (represented by a graph), directed transport network (represented by a digraph) and finally mixed transport network (represented by a migraph). For the sake of simplicity, this theoretical contribution is aimed at graph based solution approach.

Despite the obvious similarity of the both tasks, the mathematical background substantially differs. The first approach based on the vertex operation belongs to the so-called NP-hard combinatorial optimization problems and non-polynomial solutions of algorithms. Therefore, the solutions methods of these problems are time-consuming and hence unacceptable. On the contrary, the second approach based on the edge servicing is polynomial $O(n^k)$ solution, [7]. For that reason, the calculation methods of the second case are more feasible and acceptable. This paper aims to present a proposal of transformation of the assignment focusing on vertex service to the task aimed at edge service intend to reduce the computational time while taking into account an acceptable inaccuracy occurring in this transformation.

2 Problem focusing on a vertex service

The first basic step of problem solving is the task focusing on a vertex service. Let's designate it as **problem A**.

[1]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, petr.kozel.@vsb.cz
[2]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, vaclav.friedrich@vsb.cz
[3]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, sarka.michalcova@vsb.cz

2.1 Formulation of problem A

Let the connected graph is specified by $N(V, E, l)$. This one has three attributes: $V = 1, ..., m$ represents a set of vertices, $E = 1, ...n$ is a set of edges (undirected edges) and the last one l_{ij} represents an edge evaluation ij, where $i, j \in V$ is a road length (in kilometers). Example of graph is in Figure 1 and moreover, this one will be used as a model example. Generally, the challenge is to find a route passing through all vertices in graph at least once with the minimal length, but with the exception of the initial vertex, which is the endpoint simultaneously. One possible approach that can be used to solve this problem is to find the minimum Hamiltonian circuit (MHC), named by sir *William Rowan Hamilton* (1805–1865), who dealt with searching for pathways in graphs, [1].

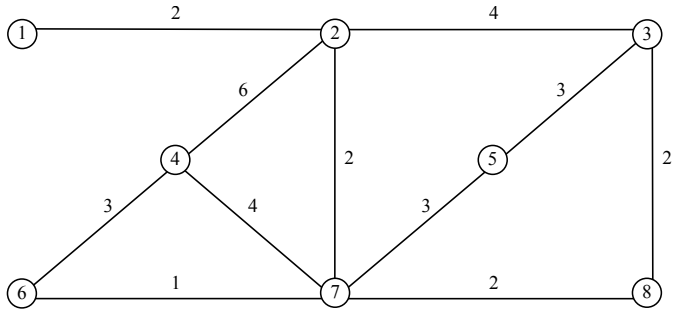


Figure 1 Diagram illustrating the graph N

2.2 Mathematical model for determining MHC

The basic condition for finding MHC is the existence of hamiltonian graph. However, graphs corresponding to the real networks often do not meet this required condition and they should be transformed into complete graphs. Distance matrix c_{ij} among individual vertices in complete graph can be calculated using a suitable algorithm (*Dijkstra's algorithm, Floyd-Warshall algorithm etc.*). Then undoubtedly, a linear mathematical model for determining MHC could be described as following, [3]:

$$\text{Min} \sum_{\substack{i=1 \\ i \neq j}}^{m} \sum_{j=1}^{m} x_{ij} \cdot c_{ij} \tag{1}$$

$$\sum_{\substack{i=1 \\ i \neq j}}^{m} x_{ij} = 1, \text{ for } j = 1, ..., m \tag{2}$$

$$\sum_{\substack{k=1 \\ j \neq k}}^{m} x_{jk} = 1, \text{ for } j = 1, ..., m \tag{3}$$

$$y_j - y_i + m \cdot x_{ij} \leq m - 1, \text{ for } i = 2, ..., m, \ j = 2, ..., m, \ i \neq j \tag{4}$$

$$x_{ij} \in \{0, 1\}, \text{ for } i = 1, ..., m, \ j = 1, ..., m, \ i \neq j \tag{5}$$

$$y_i \in \mathbb{Z}_0^+, \text{ for } i = 2, ..., m \tag{6}$$

The variable x_{ij} is a bivalent variable representing inclusion, resp. non-inclusion of the edge into the route. Objective function (1) represents the total length of route. Conditions (2) and (3) ensure that just one edge leaves one vertex and enters another. Conditions (4) prevent from creating an subcycle. Furthermore, conditions (5) and (6) determine a domain of variables x_{ij} and y_i. Admittedly, mathematical model (1) – (6) represents an exact way to determine MHC. If we apply it to the model example of the network N, we obtain solution corresponding to the sequence of vertices 1-2-7-6-4-7-8-3-5-7-2-1 and of the total length 26 km.

3 Transformation of assignment focusing on vertex service to a task aimed at edge service

Due to the fact that computational complexity of real problems using exact mathematical model (1) – (6) is considerable, it is discutable, whether this time-consuming method can be replaced by a heuristic approach, which obtains a feasible solution in a substantially shorter time. The submitted paper shows a possibility of using heuristic approach to combine following problems: *the task of determining a minimum spanning tree* (**problem B**) and *modified task to find Eulerian path* (**problem C**). To summarize, transformation of task focusing on vertex service to assignment, which is aimed at edge service can be divided into two steps:

1. determining a minimum spanning tree (MST),
2. determining "Eulerian path" (EP) on MST.

In the following, there will be introduced both tasks corresponding to the individual steps of described transformation.

3.1 Formulation of problem B

As already mentioned above, the first step of transformation is finding MST. So, there is necessary to determine MST $K(V, E_k, l)$ in graph N, which was defined in subchapter 2.1. This spanning tree is connected acyclic factor subgraph of the graph N and it contains all vertices of the graph N. Furthermore, $|E_k| = |V| - 1$ and the sum of evaluation of edges, which are included to spanning tree, is minimal.

3.2 Mathematical model for finding MST

To find MST, some of suitable algorithms should be used, for instance *Prim's algorithm, Kruskal's algorithm*. Nevertheless, the following linear mathematical model can also be used, especially for smaller-scale tasks, [2]:

$$\text{Min} \sum_{\substack{i=1 \\ i \neq j}}^{m} \sum_{j=1}^{m} x_{ij} \cdot l_{ij} \tag{7}$$

$$\sum_{\substack{i=1 \\ i \neq j}}^{m} \sum_{j=1}^{m} x_{ij} = m - 1 \tag{8}$$

$$y_j \geq y_i + 1 - m + m \cdot x_{ij}, \text{ for } i = 1, ..., m, \ j = 1, ..., m, \ i \neq j \tag{9}$$

$$\sum_{\substack{j=1 \\ i \neq j}}^{m} x_{ij} = 1, \text{ for } i = 2, ..., m \tag{10}$$

$$x_{ij} \in \{0, 1\}, \text{ for } i = 1, ..., m, \ j = 1, ..., m, \ i \neq j \tag{11}$$

$$y_i \in \mathbb{Z}_0^+, \text{ for } i = 1, ..., m \tag{12}$$

Variable x_{ij} is also a bivalent variable, which models inclusion, resp. non-inclusion of the edge ij into spanning tree. Variable l_{ij} represents the edge evaluation ij (in kilometers). A value of infinity ∞, resp. a sufficiently large prohibitive constant T is assigned to appropriate relations ij, which correspond to the absence of the edge. Condition (8) ensures that spanning tree will contain just $m - 1$ edges. In addition, conditions (9) could ensure that spanning tree will not contain any subcycles. Conditions (10) provide connection of spanning tree with initial vertex (the root of tree). And finally, conditions (11) and (12) define again a domain of variables x_{ij} and auxiliary variables y_i. If this mathematical model (7) – (12) is applied to model example mentioned above, you can obtain MST, which contains these edges: 1-2, 2-7, 3-8, 4-6, 5-7, 6-7, 7-8. The sum of the lengths of these edges is 15 km. The acquired MST is shown by dashed lines in diagram in the Figure 2.

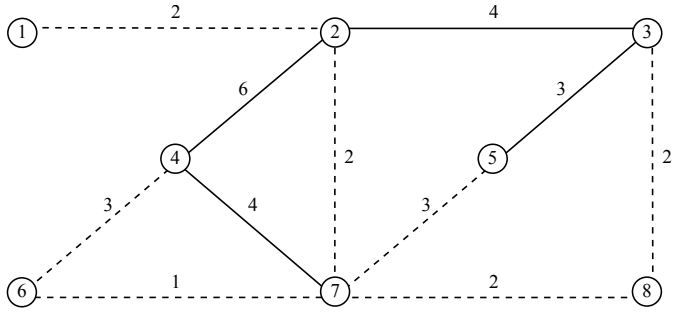


Figure 2 Diagram illustrating MST of the graph N

3.3 Formulation of problem C

The second step of transformation is finding the Eulerian path on the obtained MST. Importantly, for simple searching of the Eulerian path (*Edmonds's algorithm*) it is vital that the MST conditions of the Eulerian path are fulfilled. However, the fulfillment of conditions of existence the Eulerian path is not assumed in case of spanning tree and therefore a substitute procedure is used to minimize inevitable repetitions of some edges in the route. These approaches are discussed in detail in contributions [6], [4], in which the mathematical model has been described. The output of this model is determination of the number of transits through individual edges in graph. What is more, the paper [5] introduced another practical extension of basic approach (*modified task of finding the Eulerian path*), where edges from a wider network can be used for repeated edge transits (in practice, these are an unproductive transits of service vehicles). This approach can be applied effectively for the transformation described in just presented contribution and it will also be described in detail below.

The graph $N(V, E, l)$, which has already been described in the subchapter 2.1, can be divided into two subgraphs. The first one $N_1, (V_1, E_1, l)$ represents a set of edges, which must be included in the route. The second subgraph $N_2, (V_2, E_2, l)$ illustrates a set of edges, which can be used in the route for effective transits. The following notation between both subgraphs is valid: $N = N_1 \cup N_2$. The goal of task is to find such route, which passes through all edges of the subgraph N_1 at least once with possibility of using the edges from the subgraph N_2, so that the total length of route was minimal and that this route starts and ends at starting vertex.

If we want to use this modified task to find the Eulerian path on MST, then the subgraph N_1 will correspond to MST K. The set of vertices of subgraph N_1 is identical to the set of vertices of graph N. Admittedly $V_1 = V$, because the spanning tree contains all vertices of the relevant graph. However, the set of edges in subgraph N_1 will contain only these edges, which belong to MST, so $E_1 = E_k$. The subgraph N_2 will contain edges, which are not included to MST and which can be used for effective transits when searching the Eulerian path. The set of vertices in subgraph N_2 contains only some vertices. The set of edges in subgraph N_2 contains all other graph edges, which do not create MST K. So, $E_2 = E - E_k$.

3.4 Mathematical model for finding a number of edge transits

Mathematical model for finding a number of transits through individual edges of subgraphs N_1 and N_2 could be expressed as follows, [5]:

$$\text{Min} \sum_{\substack{i=1 \\ \text{exists } d_{ij}}}^{m} \sum_{j=1}^{m} d_{ij} \cdot z_{ij} + \sum_{\substack{i=1 \\ \text{exists } e_{ij}}}^{m} \sum_{j=1}^{m} e_{ij} \cdot y_{ij} \tag{13}$$

$$z_{ij} + z_{ji} \geq 1, \text{ for } i, j = 1, ..., m, \text{ where } \text{exists } d_{ij} \tag{14}$$

$$\sum_{\substack{j=1 \\ \text{exists } d_{ij}}}^{m} z_{ij} + \sum_{\substack{j=1 \\ \text{exists } e_{ij}}}^{m} y_{ij} = \sum_{\substack{j=1 \\ \text{exists } d_{ij}}}^{m} z_{ji} + \sum_{\substack{j=1 \\ \text{exists } e_{ij}}}^{m} y_{ji}, \text{ for } i = 1, ..., m \tag{15}$$

$$z_{ij} \in \mathbb{Z}_0^+, \text{ for } i, j = 1, ..., m, \text{ where } \text{exists } d_{ij} \tag{16}$$

$$y_{ij} \in \mathbb{Z}_0^+, \text{ for } i, j = 1, ..., m, \text{ where } \text{exists } e_{ij} \tag{17}$$

The objective function (13) indicates the total length of resulting route. The first part of expression represents the distance traveled when servicing the edges in graph N_1 (edges in MST), the second one is the distance traveled when using the edges in subgraph N_2 (edges located outside the spanning tree). Variable z_{ij} is an integer, which models a number of transits through edges in subgraph N_1. In a similar way, variable y_{ij} is an integer, which indicates a number of edge transits of subgraph N_2. Conditions (14) secure that every existing edge in subgraph N_1 will be included in route at least once. Moreover, formula (15) indicates that each served vertex will be subsequently abandoned. Obligatory conditions (16) and (17) ensure that variables z_{ij} and y_{ij} are nonnegative integers.

Mathematical model (13) – (17) can now be applied for outputs obtaining by solution of mathematical model (7) – (12). So, a set of edges E_1 of subgraph N_1 is following: 1-2, 2-7, 3-8, 4-6, 5-7, 6-7, 7-8 and a set of edges E_2 of subgraph N_2 is: 2-3, 2-4, 3-5 a 4-7 (the order of vertices defining the edge does not matter: the edge 2-3 = the edge 3-2). When the mathematical model (13) – (17) is solved, we can find out that there are all edges of subgraph N_1 in resulting route at least once. These edges corresponding to MST and two edges to subgraph N_2. Mathematical model output is summarized in Table 1. Final route is obtained by edge arranging: 1-2-4-6-7-5-3-8-7-2-1. The length of this route is 26 km and it is shown in diagram of the graph N in Figure 3. Obtained route is a closed trail, which contains all edges of subgraph N_1 and which serve all vertices in graph N at the same time. Consequently, this route could be considered as an alternative to MHC discussed in the subchapter 2.2.

z_{ij}	z_{12}	z_{72}	z_{38}	z_{46}	z_{75}	z_{67}	z_{87}	y_{ij}	y_{24}	y_{53}
	2	1	1	1	1	1	1		1	1

Table 1 Mathematical model output (13) – (17)

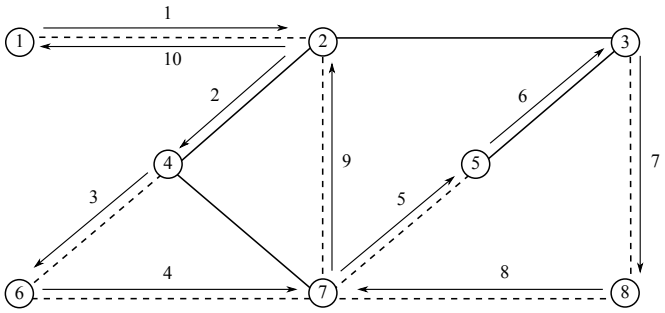


Figure 3 Diagram representing closed trail as an alternative to MHC

The value of optimal solution of MHC obtained on basis of mathematical model (1) – (6) is 26 km. Similarly, high-quality solution was obtained when applying transformation and subsequent solving mathematical models (7) – (12) and (13) – (17). This consistency is mainly due to the fact that model problem was very small-scale task. When you solve real-scale tasks, the solution of mathematical models (7) – (12) and (13) – (17) is usually worse, but with a significant difference in value of computational time. It will be discussed in more detail in next chapter dedicated to computational experiments.

4 Numerical experiments

Numerical experiments verifying designed way of transformation of the task of vertex service to the task of edge service are listed in this chapter. These experiments can be divided into three sets **A**, **B**, **C** and they are summarized in Table 2. Information of task dimension V/E (vertice/edges), the quality of obtained solution and computational time are listed at each task to find:

- minimum Hamiltonian circuit (MHC) using mathematical model (1) – (6),
- minimum spanning tree (MST) using mathematical model (7) – (12) and Kruskal's algorithm,
- Eulerian path in MST (EP) using mathematical model (13) – (17).

Furthermore, the table informs us about deterioration of the solution obtained by described transformation.

The set **A** is formed by trivial tasks, dimension of which is $m \leq 20$. The task 1/A is listed in the Table 2 with respect to similar results for all 13 tasks of the set **A**. Average time for solving the MHC task was 1.57 sec, for the MST task 0.32 sec and for the EP task 0.02 sec. Average deterioration of the solution obtained by transformation compared with the solution of basic task was 5.52%.

The set **B** contains tasks, dimension of which is $20 < m \leq 50$. The task 2/B is listed in Table 2, again with respect to similar results for all 9 tasks of the set **B**. Average time for solving the MHC task was 14920 sec, for

Exp.	V/E	MHC	MHC-time	MST	MST-time	EP	EP-time	Deterioration
	[-]/[-]	[km]	[sec]	[km]	[sec]	[km]	[sec]	[%]
1/A	20/25	73	12.8	49	0.4	77	0.1	5.48
2/B	40/59	201	1495	151	0.9	216	0.0	7.46
3/C	100/153	327	>86400	228	0.0	368	0.3	12.54
4/C	200/277	1269	>86400	886	0.0	1399	2.5	10.24
5/C	300/436	1780	>86400	1309	0.0	1989	2.4	11.74
6/C	400/546	2634	>86400	1777	0.0	2822	16.6	7.14
7/C	500/655	3471	>86400	2364	0.0	3719	7.7	7.14

Table 2 Numerical experiments

the MST task was 0.97 sec and finally for the EP task 0.11 sec. Average deterioration of the solution obtained by transformation compared with the solution of basic task was 6.39%.

At last, the set **C** contains such tasks, dimension of which is $100 \leq m \leq 500$ and solutions of all 5 tasks are presented in v Table 2.

To solve tasks of finding MHC and EP, mathematical models (1) – (6) and (13) – (17) implemented in the computational environment *Xpress-IVE* were used in all mentioned cases, [9]. Mathematical model (7) – (12) implemented in *Xpress-IVE* only for sets of tasks **A** and **B** was used for computation MST. And at last, the MST for the set **C** was found using Kruskal's algorithm in computational environment *Mathematica*, [8].

5 Conclusion

The presented paper has been dealing with the transformation of two assignments: the task focusing on vertex service to the task aimed at edge service. In general, we gradually introduced these following mathematical models: finding the minimum Hamiltonian circuit (chap. 2) and then two mathematical models focusing on realization of the explained transformation (chap. 3). Especially, computational experiments described in the chapter 4 were performed both with trivial tasks and with large dimensional real tasks. Finally, the results of experiments show that obtained solution of the transformation issue be substantially time-saving.

Acknowledgements

This paper was supported within Op. Prog. Education for Competitiveness Project No. CZ.1.07/2.3.00/20.0296 and programme for the support of applied research and experimental development, Technology Agency of the Czech Republic (EPSILON): Effective approaches to economical and adaptable systems of maintenance and operation of transport networks TH02010930.

References

[1] Applegate, D., L., Bixby, R., E., et al.: *The Traveling Salesman Problem*. Princeton University Press, New Jersey, 2006.

[2] Guéret, Ch., Prins, Ch., et al.: *Applications of optimization with Xpress-MP*. Blisworth: Dash Associates, United Kingdom, 2002.

[3] Janáček, J.: *Matematické programování*. Žilina, Žilinská univerzita, 2003.

[4] Kozel, P., Michalcová, Š.: The use of linear programming in solving service tasks of mixed transport networks. In: *Proceedings of 11th International Conference Strategic Management and its Support by Information Systems 2015*, Uherské Hradiště, 2015, 274–279.

[5] Kozel, P., Michalcová, Š.: The use of linear programming to solve routing task in practice. In: *Proceedings of 33th International Conference Mathematical Methods in Economics 2015*, Cheb, 2015, 395–400.

[6] Kozel, P., Michalcová, Š., and Friedrich, V.: The using of linear programming for solving the municipal waste collection problem. In: *Proceedings of 32nd International Conference Mathematical Methods in Economics 2014*, Olomouc, 2014, 483–488.

[7] Palúch, S.: *Algoritmická teória grafov*. Žilina, Žilinská univerzita, 2008.

[8] Wolfram, S. *The Mathematica Book*. Cambridge University Press, New York, 1999.

[9] *XPRESS-Mosel "User guide"*. Blisworth: Dash Associates, United Kingdom, 2005.

Comparison of selected DEA approaches for market risk models evaluation

Aleš Kresta[1], Tomáš Tichý[2]

Abstract. Evaluation of market risk models, typically done by means of backtesting on historical time series, is important part of risk management activities. In the paper, we compare two different approaches. We apply traditional statistical tests, namely Kupiec's unconditional coverage test (Kupiec, 1995) and its extension proposed by Pérignon and Smith (2008), and alternative approach based on data envelopment analysis proposed in Kresta and Tichý (2016). Both approaches are applied for evaluation of selected models' accuracy while assuming VaR estimation at multiple significance levels. In the practical application, we compare the normal distribution, historical simulation and Normal Inverse Gaussian distribution – a type of Lévy model. All models are backtested on daily data of Prague stock market index during the period from May 2003 until March 2016.

Keywords: backtesting, data envelopment analysis, risk management, value at risk.

JEL Classification: C44, G17
AMS Classification: 90C05, 90C90

1 Introduction

Modelling and managing of financial risks is important activity of both the banks and insurance companies. Alas, construction of a sound model for estimation of Value at Risk (VaR), which is commonly applied risk measure, is difficult.

Another issue is to evaluate the accuracy of the models, which is typically done by means of backtesting procedure on historical time series. This procedure can be simply explained as the comparison of the quantity of model failures observed over a certain period with their predicted number. Failures are here understood to be such a situations, in which the realized loss is higher than the loss predicted at a given significance level. In the case of VaR measure, these can be addressed as VaR violations. The significance level determines the expected number of failures/violations for a given horizon. For example, for a period that includes 250 observations, 5 failures/violations could be expected at a 2% significance level.

Recently, with Basel III proposals, an alternative measure called Conditional Value at Risk(CVaR), which takes into account the conditional expected loss, has been considered to replace VaR. Since VaR represents the quantile of probability distribution, which either is or is not exceeded, the backtesting of its estimation is easy. However, the evaluation of the model's accuracy of CVaR estimation is much more complicated. There are studies whose authors focus on the possibility of direct testing of CVaR risk measure, see e.g. [3], but the proposed procedures are rather demanding and not always reliable.

From the definition of VaR and CVaR it is obvious that CVaR can be approximated as the weighted mean of VaRs. Some authors thus recommend to link regulatory capital to the CvaR criterion, but to verify the model using VaR criterion. For instance, in order to backtest the CVaR, Emmer et al. [4] propose an empirical approach that consists of replacing CVaR by a set of four quantiles, which should allow us to make use of backtesting methods for VaR, i.e., in order to judge the accuracy of CVaR estimation we test the estimation of VaR at four different significance levels. However, even testing VaR estimation at multiple significance levels is not an easy task.

The aim of the paper is to compare different approaches for evaluation of backtesting results. Specifically, we compare following approaches: statistical testing approach as suggested in [9, 11] with the DEA approach as proposed in [7]. In DEA approach, we further compare the two alternative models –one with and one without weight restrictions. Moreover, we compare the results when evaluated from the point of view of a bank and its supervisor.

We proceed as follows. In the second section, we describe the Value at Risk measure. In the third section, the backtesting procedure with focus on the evaluation of its results is described. Specifically, we focus on classical

[1] VŠB - Technical University of Ostrava, Department of Finance, Sokolská tř. 33, 702 00, Ostrava, Czech Republic, ales.kresta@vsb.cz.
[2] VŠB - Technical University of Ostrava, Department of Finance, Sokolská tř. 33, 702 00, Ostrava, Czech Republic, tomas.tichy@vsb.cz.

statistical tests and alternative DEA approach. In the fourth section, we briefly describe selected data and finally, the results and comparisons are provided in the fifth section.

2 Value at risk and applicable models

Value at Risk is defined as the maximum potential loss that a portfolio can suffer within a fixed confidence level during a holding period. However, it is more convenient to work with the distribution of profits, knowing that the loss can be stated as a negative profit. Thus assuming a random-variable profit X with known cumulative distribution function F_X, VaR can be defined as follows,

$$VaR_{X,\alpha} = -\inf\left\{x \in R : F_X(x) \geq \alpha\right\}, \tag{1}$$

where α is the chosen significance level (e.g. 1%) specifying the probability with which the observed loss can exceed estimated VaR,

$$\Pr\left(-X \geq VaR_{X,\alpha}\right) = \alpha. \tag{2}$$

As it is obvious from (1), VaR is the negative value of alpha's percentile of profit probability distribution. It shall be also noted that VaR can be alternatively defined on the basis of a confidence level $1-\alpha$, i.e., the probability with which the observed loss will not exceed the VaR. Value at Risk methodology used to be strongly criticized. The criticism started with [2] by specifying the properties of coherent risk measures, which do not hold for VaR generally. Another criticism of VaR is the fact that it does not quantify the loss, which is higher than VaR.

We can generally distinguish three groups of methods for risk estimation. They are historical simulation, analytical formula and Monte Carlo simulation. These are also applied in the paper as we compare following three models for VaR estimation: normal distribution (analytical formula, henceforth Gaussian), historical simulation (henceforth HS), which was documented as by far the most popular VaR estimation method by Pérignon and Smith [12], and Normal Inverse Gaussian distribution (henceforth NIG) solved on the basis of Monte Carlo simulation. Description of the applied models can be found in our previous studies [6, 7, 8].

3 Backtesting of VaR measure for multiple significance levels

Within the backtesting procedure, the ability of a given model to estimate the future losses is tested. Backtesting is based on the estimation of the risk (mostly measured as Value at Risk) at time t for time $t+\Delta$, $VaR_{X,\alpha}(t;t+\Delta)$, and comparison with the true loss $L_X(t;t+\Delta)$ observed from time t until time $t+\Delta$. The period Δ, for which VaR is calculated, is usually (in line with the standards for bank supervision) set to one business day. Within the above-specified comparison the following two situations can arise – the loss is higher or lower than its estimation. While the former case is denoted by 1 as VaR violation (also called exception), the latter one is denoted by zero.

The whole procedure is applied on rolling window basis over the complete utilized dataset. In this way, we obtain the sequence of logical values corresponding to the fact whether the violation has or has not occurred. We get the sequence I_t of logical values,

$$I_t = \begin{cases} 1 & if\ L_X(t;t+\Delta t) > VaR_{X,\alpha}(t;t+\Delta t) \\ 0 & if\ L_X(t;t+\Delta t) \leq VaR_{X,\alpha}(t;t+\Delta t) \end{cases} \tag{3}$$

On this sequence it can be tested whether the number of ones corresponds with the assumption, i.e., $\alpha \cdot n$, where n is the length of the time series utilized for backtesting. It should be noted that n is not equal to the size of the utilized dataset as part of the data (of length m) should be left for the first risk estimation and thus cannot be utilized for backtesting. In this paper we focus solely on the quantity of VaR violations. Further, we denote the sum of VaR violations as $n_1 = \sum_{t=m+1}^{m+n} I_t$ or alternatively n_i for i-th significance level when assuming more significance levels.

3.1 Statistical approach

In this section we describe the statistical test proposed by Pérignon and Smith [11]. The test is the extension of Kupiec's [9] unconditional coverage test, thus, we briefly introduce it first. Kupiec's test is derived from a relative amount of violations, i.e., whether their quantity is from the statistical point of view different from the assumption. The null hypothesis is that the observed probability of violation occurring is equal to the significance level. A given likelihood ratio test based on chi-squared probability distribution with one degree of freedom is formulated as follows,

$$LR_{Kupiec} = -2\log\left[\frac{\alpha^{n_1}(1-\alpha)^{n-n_1}}{\left(\frac{n_1}{n}\right)^{n_1}\left(1-\frac{n_1}{n}\right)^{n-n_1}}\right], \tag{4}$$

where n_1 is the number of observed VaR violations, n is the length of backtesting period and α is the significance level. When making conclusions about the accuracy of the model based on p-value of Kupiec test, the following rule of thumb can be applied: the higher the p-value the more accurate the model.

The Kupiec's unconditional coverage test can be extended to assume multiple significance levels. Let's assume l different significance levels, α_i, i=1…l, such that $\alpha_i > \alpha_{i+1}\ \forall i$. Moreover, for each significance level we know the quantity of VaR violations, n_i, i=1…l. We can formulate the following likelihood ratio,

$$LR = -2\ln\left[\frac{(1-\alpha_1)^{n-n_1}\cdot\alpha_l^{n_l}}{\left(1-\frac{n_1}{n}\right)^{n-n_1}\cdot\left(\frac{n_l}{n}\right)^{n_l}}\cdot\frac{\prod_{i=1}^{l-1}(\alpha_i-\alpha_{i+1})^{n_i-n_{i+1}}}{\prod_{i=1}^{l-1}\left(\frac{n_i-n_{i+1}}{n}\right)^{n_i-n_{i+1}}}\right], \tag{5}$$

which has asymptotically chi-squared distribution with l degrees of freedom. In fact the tests statistics (5) is a likelihood ratio, in which we compare the probability of the null hypothesis (the probabilities of VaR violations are equal to the significance levels) to the alternative hypothesis (the probabilities of VaR violations are equal to the observed probabilities). Pérignon and Smith [11] demonstrated an illustrative example, in which by applying the extended test (5) they rejected the null hypothesis, whereas according to Kupiec's unconditional test (4) the null hypothesis cannot be rejected for any of the significance levels under consideration.

3.2 Alternative approach utilizing DEA

Data envelopment analysis (DEA) is a well-known approach, which is at present most frequently applied to measure the banks' efficiencies, see e.g. [10] and references therein. In general, it ranks the performance of decision-making units (DMU), preferably homogeneous, with multiple inputs and multiple outputs. This method, originated in [5], maximizes the ratio of the weighted sum of outputs to the weighted sum of inputs for a DMU, subject to the condition that the same ratio for all DMUs must be less than or equal one. Mathematically, suppose there are n DMUs (DMU$_j$; j= 1,…, n) with m inputs, $\mathbf{x}_j = (x_{1j}, \ldots, x_{mj})$, and s outputs, $\mathbf{y}_j = (y_{1j}, \ldots, y_{sj})$. Note that inputs and outputs are quantitative data, which are available for each DMU. The fractional programing model, which measures the relative efficiency score of evaluated DMU, i.e., DMU$_o$, o=1,…,n, can be transformed to the following linear programming problem,

$$\begin{aligned}
&\max \theta = \sum_{r=1}^{s} u_r y_{ro} \\
&\sum_{i=1}^{m} v_i x_{io} = 1 \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0 \quad j = 1,\ldots,n \\
&u_r \ge 0 \quad \forall r \\
&v_i \ge 0 \quad \forall i
\end{aligned} \tag{6}$$

This model is called CCR (adapted from Charnes, Cooper and Rhodes) and involves n+1 constraints and m+s variables (weights). Let us assume that ($\theta^*, \mathbf{u}^*, \mathbf{v}^*$) is the optimal solution to the CCR model (6). The evaluated DMU, DMU$_o$, is efficient if and only if $\theta^* = 1$ and there exists at least one strictly positive optimal solution. For practical purposes, both variables **u** and **v** might be constrained by ε, a non-Archimedean infinitesimal parameter, which is added in order to prevent the input and output weights to get zero values, see [1]. Although this restriction might be helpful for some problems, it is too loose in selection of efficient market risk models and we will assume the restriction, in which each (input) weight can be only L times higher than any other (input) weight, i.e., we add the following constraints,

$$v_i - L\cdot v_j \le 0 \quad i = 1,\ldots,m,\ j = 1,\ldots,m, \tag{7}$$

which leads to a model with restricted feasible region firstly proposed by Thompson et al.[13].

In our previous studies [6, 7, 8], we assumed that the inputs consists of differences between observed and expected quantities of VaR violations at particular significance levels while the output is the time needed to compute the VaR estimates. Due to the weight restriction (7), we modify it and as inputs we take the relative (percentage) differences instead of the absolute differences. Moreover, we compare the evaluation from point of view of

the bank (both the positive and negative differences are undesirable) and the bank's supervisor (only observed quantity higher than expected is undesirable).

4 Applied dataset

The above-described approaches are applied in market risk estimation of Prague stock market index (index PX-50 and PX, henceforth only index)[3], which consists of the most actively traded blue chips for which the liquidity is not an issue. It is capitalization-weighted price index, which is calculated in CZK and disseminated in real-time by the Prague Stock Exchange.

Series applied for backtesting is from May 2003 until March 2016, which is 3,481 daily returns (n). The prior part of the time series, i.e., from September 1993 until May 2003 (2,270 daily returns), is utilized only for the purpose of VaR estimation. Based on the data downloaded from ftp server of Prague Stock Exchange [4] we have calculated the log-returns as follows,

$$r_t = \ln(p_t) - \ln(p_{t-1}). \tag{8}$$

The evolution of index values as well as calculated log-returns are depicted in Figure 1. As it is apparent from the figure, there are present different market states in our backtesting period: i) relatively stable uptrend period (2003-2008) with the sudden correction in 2006, ii) downtrend period connected to the crisis (2008-2009) and iii) relatively stable period (2012-2016). Empirical distribution of the log-returns can be characterized in the following way: (i) although the median (0.07%) is close to the mean (0.02%), due to nonzero skewness (–0.57) we can conclude that the probability distribution of the returns is skewed, (ii) also high kurtosis (18.4) suggests the presence of heavy tails, although for particular homogenous subintervals the kurtosis would be much smaller, (iii) the autocorrelation of the returns is small, but present and (iv) the autocorrelation of the squared returns is relatively high, which suggest the presence of volatility clustering.

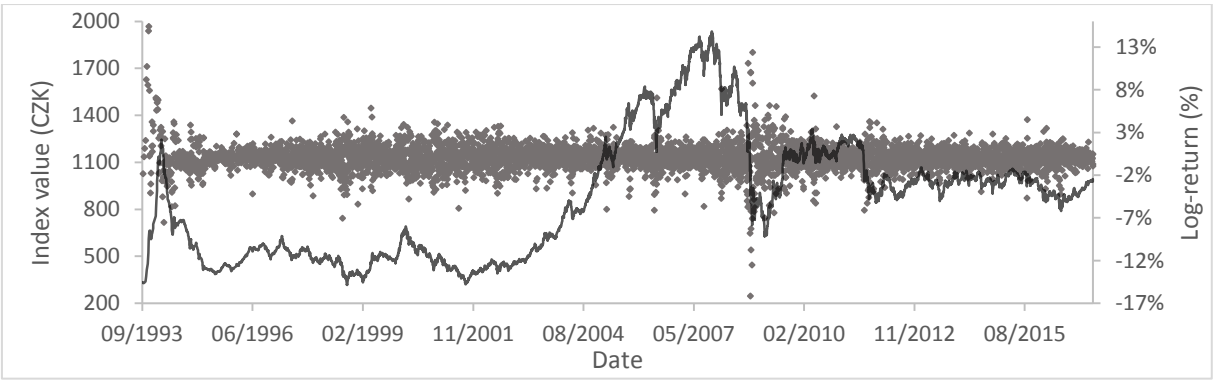


Figure 1 Daily values of the index and log-returns

5 Results

In this section, we present the evaluation of backtesting results. We compare three models, namely historical simulation (HS), normal/Gaussian distribution of returns (Gaussian) and NIG distribution of the returns (NIG). Moreover we assume different periods of parameters estimation, we compare the periods from 21 days (approximately one month) to 2,100 days (approximately eight years) with the step of 21 days. We backtested the VaR estimation for significance levels 15%, 14%, …, 1%. First, in Figure 2 we provide the results (the p-values) of statistical testing according to likelihood ratio test (5). Second, in Figure 3 we provide the results (efficiency scores) of DEA approach (6) without weight restriction and with the weight restriction (7).

Figure 2 shows the results of the statistical testing – p-values are shown in the left graph, the moving average of p-values is shown in the right graph for better clarity. As could be expected, normal distribution cannot be applied for VaR estimation – actually, for all the estimation periods the p-values are equal to zero, i.e., we can reject the null hypothesis that the observed probabilities are equal to the significance levels. When comparing HS and NIG model, we can conclude that NIG model has on average slightly higher p-values, which signals better accuracy. There are 18 periods, in which the accuracy of HS can be rejected (p-values smaller than 5%). The model is not accurate for short estimation periods (18-84 days), which could be expected, and for long periods (1512-1827 days). For NIG model, it is not accurate only for short periods of estimation (21-105, 5 periods) and totally

[3] The index PX is calculated from March 20, 2006. In that date, it took over the values of the index PX-50, which was calculated from April 5, 1994. The previous values of index PX 50 has been calculated ex-post.
[4] http://ftp.pse.cz/Info.bas/Cz/PX.csv

there are only 8 estimation periods for which the null hypothesis can be rejected. Note that, from statistical point of view, due to the type I error, there are on average 5 periods, in which the null hypothesis is wrongly rejected.

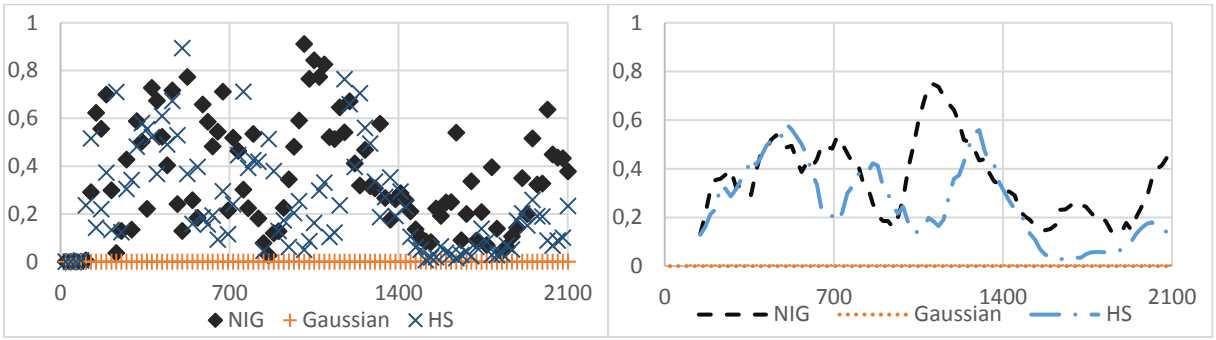


Figure 2 P-values of the statistical test

As it is obvious from Figure 2, for both NIG and HS models the null hypothesis cannot be rejected for almost all estimation periods. Even if we increase the p-value level, there would still be plenty of periods for which the models' accuracy cannot be rejected. However, we need to choose one model, which is the best. The best model would be the one for which the observed quantities would be closest to the expected. Alas, there is no model/period, which would dominate all the others (i.e., the difference between observed and expected numbers is the same or smaller than in other models/periods for each considered significance level).

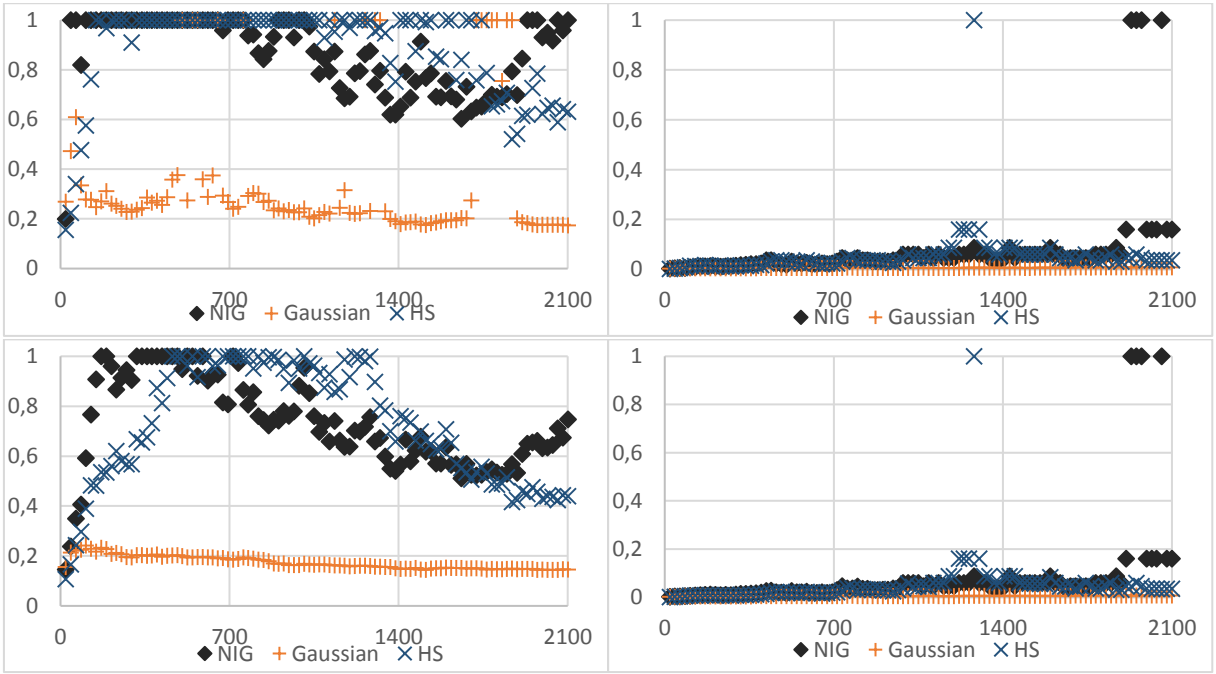


Figure 3 Efficiency scores under particular models/approaches

In order to find the efficient models/periods, we apply the DEA approach as described in section 3.2. Actually, we examine four different approaches in which the inputs are always the relative (percentage) differences between observed quantity of VaR violations and their expectations. The approaches differ in the following ways.

- The bank's point of view – the model is unwanted to overestimate the risk (observed quantity is lower), which would increase the capital requirements, as well as to underestimate the risk (observed quantity is higher). We take the absolute values of differences as the inputs. Results are presented in the left graphs in Figure 3.
- Regulator's point of view – overestimation of the risk is not undesirable, only underestimation. The negative values of differences are replaced by zero. The results are presented in the right graphs in Figure 3.
- We apply the model without weight restriction, i.e., we solve the problem (6) without the constraints (7). The results are presented in the top graphs in Figure 3.
- We restrict the weights so that any weight cannot be more than four times higher than the others, i.e., we solve problem (6) with the constraints (7). The results are presented in the bottom graphs in Figure 3.

According to the results in Figure 3 the following findings should be mentioned. The normal distribution should not be applied – the efficiency scores are around 0.2 except for some periods under unrestricted DEA model, however, these are occasional and due to unrestricted weights (at one significance level, usually 5% as found in our previous studies, the difference can be the lowest one). Applying weight-restricted DEA approach less models/periods are found as efficient, however, only in the banks's point of view. For the regulators' point of view there is no difference between unrestricted and restricted weights approach – we observe that for longer periods the risk is generally overestimated and the longer the period the more significance levels for which the difference is negative (zero). The efficient model from regulator's point of view is NIG model with long period (around 2,000 days, i.e., around 8 years), which, however, overestimates the risk and thus is not efficient from the bank's point of view (see the left graphs in the Figure 3). Considering the negative differences as serious as positive, the best models would be the NIG model with the estimation period 42-756 days and HS with the estimation period 147-1,071 days. For these models/periods, also the accuracy of the model cannot be rejected by means of statistical test (see Figure 2).

6 Conclusion

For financial institutions, modelling and managing of financial risks is important activity. Alas, the construction of sound model for estimation of Value at Risk, which is commonly applied risk measure, is difficult as well as the evaluation of its accuracy, which is typically done by means of backtesting procedure on historical time series. Traditional approach is based on the statistical testing, which may lead to a situation, in which many models are statistically accepted and the subsequent selection of the most appropriate model is problematic.

In our paper, we further developed a selection approach of a suitable model based on the DEA. We have found out expected results such as inapplicability of normal distribution and less models marked as efficient in the restricted DEA approach. Moreover, we have found out that there is the difference between banks' point of view (the overestimation of the risk is as serious as underestimation) and supervisor's point of view (the overestimation of the risk is not undesirable). We also suggest to focus on detailed evaluation of further Lévy type models, since they can improve the risk estimation substantially.

Acknowledgements

The research was supported by an SGS project of VSB-TU Ostrava under No. SP2017/32 and the Czech Science Foundation through project No. 13-13142S. The support is greatly acknowledged.

References

[1] Amin, G.R. and Toloo, M.: A polynomial-time algorithm for finding ε in DEA models. *Computers & Operations Research* **31** (2004), 803–805.
[2] Artzner, P., et al.: Coherent measures of risk. *Mathematical Finance* **9** (1999), 203–228.
[3] Du, Z. and Escanciano, J.C.: Backtesting Expected Shortfall: Accounting for Tail Risk. *Management Science* **63** (2016), 940–958.
[4] Emmer, S., Kratz, M. and Tasche, D.: What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk* **18** (2015), 31–60.
[5] Charnes, A., Cooper, W.W. and Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.
[6] Kresta, A., Tichý, T. and Toloo, M.: Posouzení modelů odhadu tržního rizika s využitím DEA přístupu. *Politická ekonomie* **65** (2017), 161–178.
[7] Kresta, A. and Tichý, T.: Selection of efficient market risk models: Backtesting results evaluation with DEA approach. *Computers & Industrial Engineering* **102** (2016), 331–339.
[8] Kresta, A., Toloo, M. and Tichý, T.: Evaluation of market risk models: changing the weights in the DEA approach. In: *34th International Conference Mathematical Methods in Economics (MME) : proceedings papers*. Technical University of Liberec, Liberec, 2016.
[9] Kupiec, P.H.: Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* **3** (1995), 73–83.
[10] Palečková, I.: Efficiency Change of Banking Sectors and Banks in the Financial Conglomerates in Visegrad Group Countries. *Ekonomicky časopis* **65** (2017), 79–92.
[11] Perignon, C. and Smith, D.R.: A new approach to comparing VaR estimation methods. *The Journal of Derivatives* **16** (2008), 54–66.
[12] Pérignon, C. and Smith, D.R.: The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking & Finance* **34** (2010), 362–377.
[13] Thompson, R.G., et al.: Comparative site evaluations for locating a high-energy physics lab in Texas. *Interfaces* **16** (1986), 35–49.

Herding, minority game, market clearing and efficient markets in a simple spin model framework

Ladislav Kristoufek[1], Miloslav Vosvrda[2]

Abstract. We present a novel approach towards the financial Ising model. Most studies utilize the model to find settings which generate returns closely mimicking the financial stylized facts such as fat tails, volatility clustering and persistence, and others. We tackle the model utility from the other side and look for the combination of parameters which yields return dynamics of the efficient market in the view of the efficient market hypothesis. Working with the Ising model, we are able to present nicely interpretable results as the model is based on only two parameters. Apart from showing the results of our simulation study, we offer a new interpretation of the Ising model parameters via inverse temperature and entropy. We show that in fact market frictions (to a certain level) and herding behavior of the market participants do not go against market efficiency but what is more, they are needed for the markets to be efficient.

Keywords: Ising model, efficient market hypothesis, Monte Carlo simulation

JEL classification: G02, G14, G17

1 Introduction

Agent-based models (ABM) have attracted much attention in economics and finance in recent years [1, 2, 3] as they describe the reality better than simplified models of traditional economics and finance. In finance, the founding contributions were laid by Brock and Hommes models [4, 5] characteristic by strategy-switching agents and possible bifurcation dynamics. Essential contributions to the topic are the early papers of [6] and [7] who introduce a possibility of generating the returns-like series from simple models based on interactions between multiple agents. They both serve as a starting point to an important branch of the ABMs which is based on a parallel between ferromagnetism and market dynamics, i.e. the Ising model adjusted for financial markets. In the models, economic agents participating in the market are spins of a magnet. In the same way as the spins, the agents are influenced by (make their decisions based on) their neighbors, or agents with similar beliefs, but also by the overall market sentiment and activity. The novel model of [8] combining the standard Ising model from physics with an additional term reminiscent of the minority game, i.e. the tendency of agents of leaning away from the majority opinion when the majority prevails too much, has been shown to successfully mimic the basic financial stylized facts such as no serial correlation of returns, persistence and clustering of volatility, and non-Gaussian distribution of returns. [9] expand the model of [8] by four additional parameters to allow for simulations of the traded volume through the balance between supply and demand. Implications for bull and bear markets together with bubbles occurrence is discussed there as well. These founding papers have led to various adjusted and generalized models trying mainly to fit the market data or mimic the stylized facts. For a detailed treatment and history of the Ising-type models in financial economics together with other agent-based models, we suggest the current treatment of the topic by [10].

We contribute to the topical literature by inspecting the implications of the financial Ising model towards the capital markets efficiency. We focus on the model parameters and how they influence returns dynamics in the optics of the efficient market hypothesis. The attention is given to finding a combination of parameters which yields an efficient market or dynamics close to it. We thus regard the question "What combination of parameters yields returns and volatility mimicking the stylized facts?" as studied and answered in enough detail in the reviewed papers, implying that the structure and construction of the models are reasonable, and we focus on the question "What combination of parameters yields returns consistent with the efficient market hypothesis?". We show that the effects of parameters are more complicated than one might expect and their influence is apparently non-linear with a special role of the critical temperature of the system and we discuss the implications for foundations of the efficient market hypothesis.

[1]Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodarenskou Vezi 4, 182 08, Prague, Czech Republic, EU, and Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00, Prague, Czech Republic, EU, kristouf@utia.cas.cz

[2]Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodarenskou Vezi 4, 182 08, Prague, Czech Republic, EU, and Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00, Prague, Czech Republic, EU, vosvrda@utia.cas.cz

2 Ising model for financial markets

As a representative of the agent-based models applied to finance and financial economics, we opt for a simple Ising model adjusted for financial markets as proposed by [8]. There are two main reasons why this specific model is chosen. First, the model is able to mimic the most important stylized facts of financial returns. And second, the model has only two parameters which allows for a straightforward interpretation of the outcomes without a need for additional restrictions.

The model builds on a combination of the standard Ising model of ferromagnetism with local field interactions [11] and a minority game behavior of market agents [12, 13]. Financial market is represented by a square lattice (usually with torus-like neighborhoods) with a side of N, i.e. with N^2 elements representing market agents. These elements are referred to as spins due to their magnetization of either $+1$ or -1. This spin orientation is translated into a financial market as either a buy or a sell signal (decision), respectively. The spin orientation of agent i for a time period t is labelled as $S_i(t)$. For each agent i, the local field $h_i(t)$ for a time period t is defined as

$$h_i(t) = \sum_{j=1}^{N^2} J_{ij} S_j(t) - \alpha C_i(t) \frac{1}{N^2} \sum_{j=1}^{N^2} S_j(t), \tag{1}$$

where $t = 1, \ldots, T$ is discrete time. The first term is defined as a local Ising Hamiltonian with neighbor interactions J_{ij}. This is the reference to the standard Ising model. In the economic interpretation, this represents the potential herding behavior as agents are influenced by their closest neighbors and they might thus tend together, potentially forming speculative bubbles. The second term represents the global coupling as it depends on the total magnetization of the system $M(t) \equiv \frac{1}{N^2} \sum_{j=1}^{N^2} S_j(t)$ at time t with sensitivity α. From the economic perspective, this term is a built-in minority game. For $\alpha > 0$, there is a tendency to go against the overall magnetization and thus against the whole market dynamics. The strategy spin $C_i(t)$ allows for deviations from the minority game behavior of spin i, i.e. $C_i(t)$ is not necessarily equal to one. On the one hand, $C_i = -1$ implies that the agents align with the total magnetization so that they follow the market trend. Such agents are usually referred to as the trend followers or chartists. On the other hand, $C_i = 1$ suggests the minority game behavior of the agents as they oppose the sign of the total magnetization. These agents are standardly referred to as the fundamentalists.

The price and returns dynamics of the system is extracted directly from the magnetization dynamics so that

$$\begin{aligned} \log P(t) = M(t) &\equiv \frac{1}{N^2} \sum_{j=1}^{N^2} S_j(t), \\ r(t) = \Delta \log P(t) &= \Delta M(t) = M(t) - M(t-1). \end{aligned} \tag{2}$$

The logic behind such representation is based on taking the positive spins as demand and the negative spins as supply. Their sum, i.e. the difference between demand and supply, is taken as excess demand so that the difference between two consecutive excess demands is a change in price of an asset [8, 9, 14].

Orientation of the spin i at time $t+1$ is given by the heat-bath dynamics transition function with probability $p_i(t)$ as

$$\begin{aligned} S_i(t+1) = +1 \text{ with } p_i(t) &= [1 + \exp(-2\beta h_i(t))]^{-1} \\ S_i(t+1) = -1 \text{ with } 1 - p_i(t), & \end{aligned} \tag{3}$$

which is directly connected to Eq. 1 with an additional sensitivity β, which is parallel to the inverse temperature of the original Ising model, i.e. $\beta = \frac{1}{T}$, and it is essential as it controls the responsiveness of the spin change probability to the local field $h_i(t)$. The inverse temperature determines the system regime – either paramagnetic or ferromagnetic in the original Ising model terminology. These two types of behavior can be easily inferred from the heat-bath dynamics in Eq. 3. For the paramagnetic regime, the transition function is rather flat so that the spin probability depends on the local field $h_i(t)$ only weakly. Decreasing β then leads to a weakening local interactions effect. For β close to zero (infinite temperatures), the spin change is completely random with probability $\frac{1}{2}$. For the ferromagnetic regime, the local interactions become more dominant forming large clusters of oriented spins, one of which eventually dominates and leads to the stable state of the model with $|M(t)| \approx 1$. [15] show that for the original Ising model, i.e. with $\alpha = 0$, the critical temperature is equal to $T_C = \frac{2}{\ln(1+\sqrt{2})} \approx 2.269$ which gives $\beta_C = \frac{1}{T_C} \approx 0.441$.

The strategy term $C_i(t)$ is given as a general term in Eq. 1 which can be further specified. A popular choice is to highlight the minority game behavior of the spin by allowing the strategy to change with respect to the total magnetization and the spin's own orientation. This specification also allows for more strategy types. [8] proposes

the following dynamics

$$C_i(t+1) = -C_i(t) \text{ if } \alpha S_i(t) C_i(t) \sum_{j=1}^{N^2} S_j(t) < 0 \tag{4}$$

for $i = 1, \ldots, N^2$. The practical implications of such rule are the following. The second term of the local field (Eq. 1) of all majority agents (who are $C_i(t) = 1$) has an opposite sign compared to the total magnetization. This forces the agents to swap their strategy. Similarly for the minority agents (with $C_i(t) = -1$), the second term of the local field has the same sign as the total magnetization, which forces them to change their strategy as well. As the total magnetization $M(t)$ is a part of the second term of the local field, the tendency towards switching strategies strengthens with the total magnetization deviating from zero, which is parallel to the equilibrium price of the asset. The further the magnetization (price) deviates from zero (equilibrium) the more agents will oppose it. In practice, this protects the model from deviating towards ± 1 and stabilizing there while still remaining well in the logic of how the market works and how the agents behave.

A simple alternative is to have the strategy spin update immediately, which reduces the local field equation to

$$h_i(t) = \sum_{j=1}^{N^2} J_{ij} S_j(t) - \alpha S_i(t) \left| \frac{1}{N^2} \sum_{j=1}^{N^2} S_j(t) \right|, \tag{5}$$

for $i = 1, \ldots, N^2$, i.e. it does not depend on the strategy of any spin at all [8]. The second term thus motivates an agent to change its spin orientation (i.e. the minority game behavior) with an increasing absolute value of magmetization $|M(t)|$.

3 Efficient market hypothesis

The efficient market hypothesis (EMH) has been a cornerstone of modern financial economics for decades. Even though its validity has been challenged on many fronts, it still remains the firm theoretical basis of the financial economics theory [16, 17]. In the fundamental paper, [18] summarizes the empirical validations of the theoretical papers of [19] and [20]. The theory is revised and made clearer in [21].

From mathematical standpoint, the historical papers [19, 20] are more important as they provide specific model forms of an efficient market. Specifically, [19] connects the (logarithmic) price process of an efficient market to a random walk and [20] specifies it as a martingale. Implications for the statistical properties of the (logarithmic) returns process, i.e. the first differences of (logarithmic) prices, of the efficient market are straightforward. For the former, the (logarithmic) returns are expected to be serially uncorrelated and follow the Gaussian (normal) distribution, i.e. the dynamics follows the Gaussian noise, which implies serial independence. For the latter, only the serial uncorrelatedness is implied as the martingale difference process is expected for the (logarithmic) returns [22, 23]. We thus have two straightforward implications of the market efficiency – (asymptotically) normally distributed (for the random walk definition) and serially uncorrelated (for both random walk and martingale definitions) returns as serial independence implies no serial correlation.

We are interested in the ability of the Ising model defined between Eqs. 1-5 to meet the criteria attributed to the efficient capital market, i.e. normality and serial uncorrelatedness of returns. To test these, we use the Jarque-Bera test [24] and Ljung-Box test [25], respectively. There are two crucial parameters in the model – α and β – which can influence the prices and returns dynamics emerging from the model. We vary these two parameters and study how they influence the rejection rate of normality and uncorrelatedness of the respective tests. In other words, we are interested in a proportion of times these tests reject (with a significance level of 0.90) market efficiency of series generated by the financial Ising model with the specified parameters. Based on findings of previous research [8], we manipulate α between 0 and 15 with a step of 1 and β between 0 and 4 with a step of 0.25. We fix the time series length to $T = 1000$ and the number of agents in the market to $N^2 = 25^2 = 625$. The neighborhood influence J_{ij} is set equal to 1 for the nearest neighbors and the spin's own position (five spins in total), and 0 otherwise. For each setting, we perform 222 simulations. Two specifications are studied – Model I given by Eq. 5, i.e. with the instant strategy spin decision, and Model II given by Eq. 1, i.e. with the standard variable strategy spins.

4 Results and discussion

We examine the effect of different combinations of parameters α and β on the returns generated by the Ising model, namely its two local field specifications given by Eq. 5 (Model I) and Eq. 1 (Model II). The former model is a simplified version which attributes a global minority game behavior to all agents (for $\alpha > 0$) whereas the latter one allows the agents to switch their global strategy between the minority game and trend following. Both

models keep their local interactions so that their decision is influenced by their nearest neighbors (for $\beta > 0$). For both models and their specifications given by the parameter setting, we run 222 simulations and for each, we test whether the generated returns are serially correlated and distributed according to the Gaussian distribution. Fig. 1 (top) illustrates the results for the "no autocorrelation" null hypothesis of the Ljung-Box test and Fig. 1 (bottom) shows the results for the "Gaussian" null hypothesis of the Jarque-Bera test for both models. For both tests, we present the rejection rate of the test with a significance level set to 90%, i.e. the proportion of simulations which generate returns inconsistent with the efficient market hypothesis. The lower the rate (or rather the closer the rate to 0.1), the closer the model specification simulates the efficient market (with respect to either serial uncorrelatedness or normality of returns).

As the no serial correlation condition is common for both specifications of the market efficiency, we start with its results. Fig. 1 (top) summarizes the simulation results for both models as heat maps and contour plots for better visualization. We observe that the outcomes are qualitatively very similar for both models. The plane is practically split into two which are separated by $\beta = 0.5$. Note that this value is close to the critical inverse temperature $\beta_C \approx 0.441$. In both parts, we find a strongly non-linear dependence between β and the rejection rate. For the models above the critical temperature (below the critical inverse temperature), we find the minimum rejection rate of around 0.5 for $\beta = 0.25$. For the models below the critical temperature (above the critical inverse temperature), the minimum rejection rate of around 0.4 is found for $\beta = 1.25$. For specifications where the local fields plays no role ($\beta = 0$), the null hypothesis of no serial correlation (and thus the efficient market hypothesis) is rejected in practically all the cases. The dependence of the rejection rate on α is much more straightforward as the higher the α parameter is, the higher the rejection rate is as well. Even though the relationship is not linear either, it is monotone. Situations closest to the efficient capital market are thus found for $\alpha = 0$. The rejection rates of the no autocorrelation hypothesis are in general higher for Model II, i.e. the model with more heterogeneous agents able to switch their strategy spin.

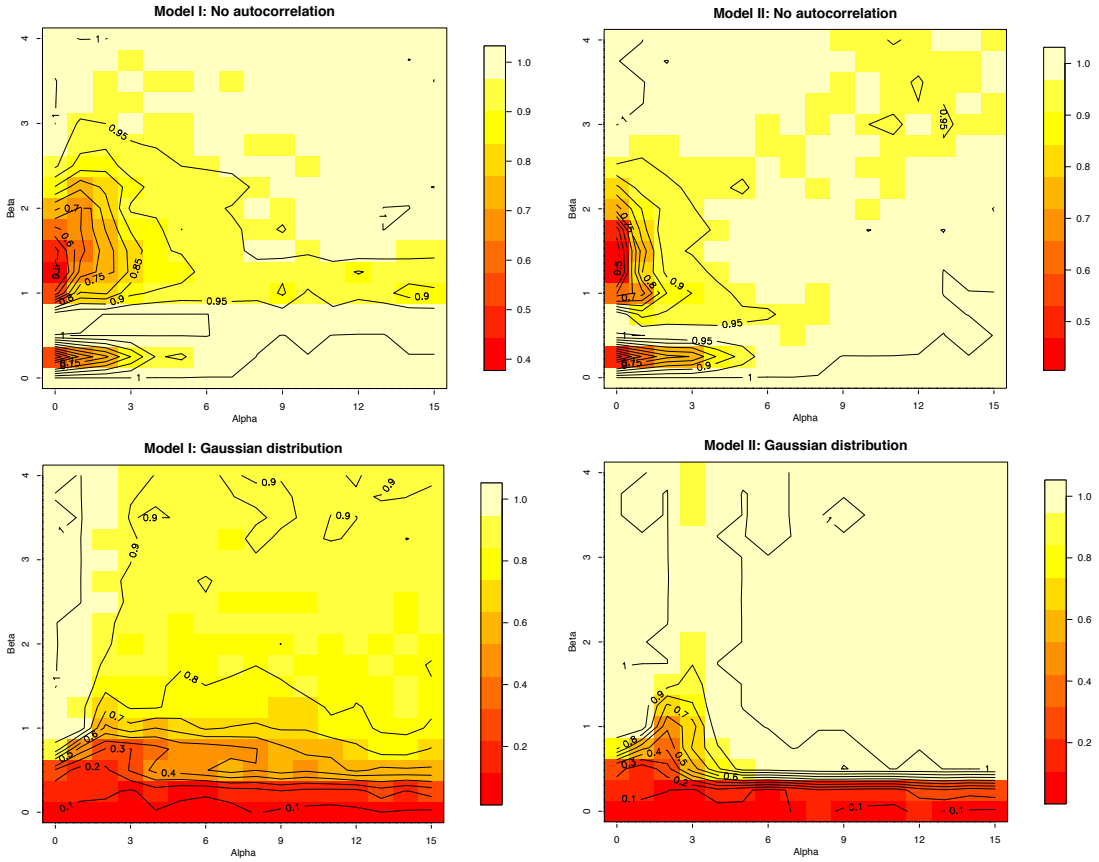


Figure 1 Rejection rates of no serial correlation (top) and Gaussian distribution (bottom) hypotheses for Model I according to Eq. 5 and Model II according to Eqs. 1 and 4.

The results for rejection rates of the Gaussian distribution are much less complex. In Fig. 1 (bottom), we find that the rejection rates attain low values only for the inverse temperatures β below the critical value. Above the critical inverse temperature β_C, the rejection rates quickly jump towards high values. This is true for both specifications of the Ising model analyzed here, even though the rejection rates are again lower for Model I. These

findings are only mildly dependent on the global coupling parameter α. For Model I, the rejection rates form a weak U-shape, i.e. the rejection rates are the highest for very low and very high levels of α, and they remain relatively lower in between. For Model II, the lowest levels are obtained for $1.5 \leq \alpha \leq 3$.

The results suggest that the model is able to generate serially uncorrelated and normally distributed returns only for a rather narrow range of parameters. Interestingly, the serially uncorrelated returns are found also for $\beta > \beta_C$ which is a new finding not discussed in the literature which usually focuses only on $\beta < \beta_C$. The model dynamics for the inverse temperatures above the critical value is thus not as uninteresting as usually claimed. From the perspective of the Gaussian distribution, though, the inverse temperatures above the critical one are not interesting.

Let us now focus on the results through the optics of the efficient market hypothesis. If we focus on the martingale version of the hypothesis, we are interested only in the serial correlation of returns. For these, we find the minimum rejection rates at $\{\alpha, \beta\} = \{0, 0.25\}$ and $\{\alpha, \beta\} = \{0, 1.25\}$. If we stick to the classical interpretation of α and β as the intensities of the global and the local coupling, respectively, we can argue that the efficient market is found for no global coupling but some local coupling. The latter part of the claim is very interesting as it suggests that some form of herding is necessary for the market to be efficient. For no local coupling with $\beta = 0$, the market is identified as inefficient for all values of α practically always. This is well in hand with an intuitive feeling that markets would not work if they were completely random, which would be the case for $\{\alpha, \beta\} = \{0, 0\}$ when the agents make their decisions on the 50-50 basis. The effect of the global coupling is rather intuitive as well – the stronger the tendency towards the minority game behavior, the lower the efficiency. The slight differences between Model I and Model II suggest that the higher heterogeneity of the agents leads to lower efficiency. When we add the Gaussian distribution into the mix, the situations when $\beta > \beta_C$ are discarded and we are left only with the under-critical inverse temperatures which are consistent with the efficient market. The interpretation as presented above is not touched by this change.

However, there is an alternative way how to interpret the interplay between α and β. Going back to the definition of the local field in Eq. 1 and the buy-sell decision probability in Eq. 3, we observe that the β parameter is present only in Eq. 3 and not in Eq. 1. Its interpretation as the intensity of local coupling (herding) is thus rather far-fetched. If we take the local field definition as an interaction between the local (first term) and global (second term) coupling, then the α parameter becomes a weight of how much more important the global coupling is compared to the local one. The higher the α parameter is, the more influence the global coupling compared to the local coupling has. If $\alpha = 0$, the dynamics is driven solely by the local coupling, and if $\alpha \gg 1$, the dynamics is driven solely by the global coupling. The fact that the generated returns are closer to the efficient market for low values of α underlines that some level of local interactions goes well in hand with market efficiency. The high values of α and thus high influence of global coupling goes directly against the market efficiency.

Such interpretations are not much different from the ones made using the standard interpretation of α and β. However, we are able to make such claims using only one of the parameters. To look deeper into the interpretation of β, we use the idea presented in [14] who discusses the market efficiency in the sense of market clearing, i.e. clearing of supply and demand, and its connection to entropy of the market. We will refer to this type of efficiency as the technical efficiency of the market. If market clears perfectly, it is technically efficient. Going back to the parallel of the original Ising model towards financial applications, we further explore possible connections between the physics model and its financial application. In practice, 100% efficiency is impossible. However, the efficient market hypothesis assumes perfect market clearing and thus the 100% efficiency. Such level of efficiency suggests that there is no energy loss in the system and as such, the entropy of the system does not increase (in general, it either increases or keeps its level). If there is energy coming into the system (i.e. agents take actions), it is only possible to have no change in entropy if the temperature of the system approaches infinity. This yields zero inverse temperature β. For positive inverse temperatures, the system entropy increases and it is not technically efficient. This gives us a new interpretation of the β parameter in the financial Ising model.

The results clearly show that the markets are not efficient in the EMH perspective for $\beta = 0$ which is parallel to the perfect market clearing. This suggests that at least some market frictions are necessary for the market to be efficient. Note that such claim does not go completely against the notion of the market efficiency as laid down by [18] who states three sufficient conditions for efficient markets – no transaction costs, all available information freely available to all agents, and all agents agree on implications of such information and future distributions of the traded assets. Fama refers to these conditions as "a frictionless market". However, he adds that these are sufficient but not necessary conditions. As specifically noted by [18], such assumptions do not reflect the real financial markets. Violating these assumptions does not necessarily imply inefficiency but it is a potential source. Our results suggest that not only the frictions do not always go against efficiency, but they mainly suggest that frictions are needed for the market to be efficient in the EMH sense. To reach the efficient market, there need to be frictions.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. FP7-SSH-612955 (FinMaP) and the Czech Science Foundation project No. P402/12/G097 "DYME – Dynamic Models in Economics". Special thanks go to Pavel Dvorak, Stepan Chrz and Pavel Irinkov for their helpful insights.

References

[1] Cars H. Hommes. Chapter 23: Heterogeneous agent models in economics and finance. In L. Tesfatsion and K.L. Judd, editors, *Handbook of Computational Economics*, volume 2 of *Handbook of Computational Economics*, pages 1109–1186. Elsevier, 2006.

[2] Blake LeBaron and Leigh Tesfatsion. Modeling macroeconomies as open-ended dynamic systems of interacting agents. *The American Economic Review*, 98(2):246–250, 2008.

[3] Joseph E. Stiglitz and Mauro Gallegati. Heterogeneous interacting agent models for understanding monetary economies. *Eastern Economic Journal*, 37:6–12, 2011.

[4] W. A. Brock and C. H. Hommes. A rational route to randomness. *Econometrica*, 65(5):1059–1095, 1997.

[5] W. A. Brock and C. H. Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics & Control*, 22:1235–1274, 1998.

[6] T. Lux and M. Marchesi. Scaling and criticality in stochastic multi-agent model of financial market. *Nature*, 397:498–500, 1999.

[7] T. Kaizoji. Speculative bubbles and crashes in stock markets: An interacting-agent model of speculative activity. *Physica A*, 287:493–506, 2000.

[8] S. Bornholdt. Expectation bubbles in a spin model of markets: Intermittency from frustration across scales. *International Journal of Modern Physics C*, 12(5):667–674, 2001.

[9] T. Kaizoji, S. Bornholdt, and Y. Fujiwara. Ddynamic of price and trading volume in a spin model of stock markets with heterogeneous agents. *Physica A*, 316:441–452, 2002.

[10] D. Sornette. Physics and financial economics (1774-2014): Puzzles, ising and agent-based models. *Reports on Progress in Physics*, 77:062001, 2014.

[11] E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift fur Physik*, 31(1):253–258, 1925.

[12] W.B. Arthur. Inductive reasoning and bounded rationality. *American Economic Review*, 84:406–411, 1994.

[13] D. Challet and Y.-C. Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A*, 246:407–418, 1997.

[14] J.L. McCauley. *Dynamics of markets: The new financial economics*. Cambridge University Press, 2009.

[15] H. Kramers and G. Wannier. Statistics for the two-dimensional ferromagnet. Part II. *Physical Review*, 60(3):263–276, 1941.

[16] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[17] B. Malkiel. The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1):59–82, 2003.

[18] E. Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25:383–417, 1970.

[19] E. Fama. The behavior of stock market prices. *Journal of Business*, 38:34–105, 1965.

[20] P. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6:41–49, 1965.

[21] E. Fama. Efficient Capital Markets: II. *Journal of Finance*, 46(5):1575–1617, 1991.

[22] J.D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[23] J. Davidson. *Stochastic Limit Theory*. Oxford University Press, 1994.

[24] C. Jarque and A. Bera. Efficient tests for normality, homoskedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters*, 7(4):313–318, 1981.

[25] G.M. Ljung and G.E.P. Box. On a measure of a lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

The Role of Credit Standards as an Indicator of the Supply of Credit

Zuzana Kučerová[1], Jitka Poměnková[2]

Abstract. We focus on the credit standards for approving loans to enterprises and households of the Euro Area banks as an indicator of the supply of credit to answer a question whether the change of credit standards is in accordance with the loan supply. The aim of the paper is to assess the co-movement between bank loans and credit standards in Euro Area via moving and dynamic correlation and also from the wavelet co-spectrum perspectives. We use quarterly data over the period 2003/Q1-2016/Q1. We conclude that banks originated fewer loans after they tightened lending standards. However, the co-movement was more apparent in case of households when the level of new loans provided by banks was lower after the lending standards tightening (and vice versa). In case of enterprises, the behaviour of banks concerning credit standards and the loans to enterprises was more unstable. In both cases, the co-movement is characterised by short-term reactions rather than long-term trend. However, the reactions changed in two years after the financial crisis of 2007 and then in 2011when banks started following their credit standards and limited the level of new loans.

Key words: Wavelet co-spectrum, comovement, bank loans, credit standards.

AMS Classification: 62P20, 62M10, 91B84

1 Introduction

Financial crisis of 2007 revealed the pro-cyclical character of the bank behaviour, i.e. that banks provide loans to clients in the growing phase of the economic cycle, however, they change the behaviour in the presence of the financial crisis when they reassess the riskiness of clients and strongly limit the supply of new loans and as such they may force the financial crises and then trigger economic crisis. Financial system of the EU countries can be characterised as a bank-based financial system (opposed to market-based financial system in the US), i.e. economic agents cannot easily substitute between a bank credit and other forms of financing in financial markets. Thus, lending capacity of banks can affect the activities of enterprises or the consumption of households dependent on bank loans. Therefore, European enterprises or households could become dependent on bank intermediation and as such banks could increase financial distress in the economy by the restriction of the loan supply. This behaviour can subsequently have an impact on the real economy (to limit the economic growth and employment). One way how to measure this behaviour of banks is to collect information about credit standards for approving new loans and to use it as an indicator of this behaviour and then try to measure the co-movement of credit standards and the volume of new bank loans.

Measuring the mutual dependence and co-movement between time series is in the forefront of many scientists' interest. The selection of appropriate approach is motivated by the application area (economy, sociology, biology, engineering etc.). In recent decades, there has been a greater degree of interdependence between disciplines and different techniques which were predominantly used in one area. In such way, economy mostly used an analysis in the time domain. From the co-movement perspective, the standard approach covered correlation techniques in classic or moving form and similar methods. Moreover, the regression analysis in the simple, the multivariate or the rolling forms can be used (Korhonen and Peresetsky [14]). Also its improvement into the vector autoregressive model (Feldkircher and Korhonen [6]), correction models (Engle and Granger [3]) or the cross-sectional dependence and the cointegration models (Hassan *et al*. [10]) can be used.

All of these techniques try to answer the question, how the synchronization of time series behaves in time and how temporal such behaviour is. One group of analysts investigate temporal character of measurement, such as dynamic conditional correlation (Lukmanova and Tondl [16]), where the dynamic behaviour is based on information known from the previous period. Lu *et al*. [17] propose the time-varying coefficient vector autoregression model, in which the coefficient is a linear function of dynamic lagged correlation. The second group

[1] Brno University of Technology, Faculty of Electrical Engineering and Communication, Technická 12, 616 00 Brno. Mendel University in Brno, Faculty of Business and Economics, Zemědělská 1, 613 00 Brno, Czech Republic.

[2] Brno University of Technology, Faculty of Electrical Engineering and Communication, Technická 12, 616 00 Brno, Czech Republic, e-mail: pomenkaj@feec.vutbr.cz.

of analysts try to answer the question using the techniques in frequency domain, which also take into account the frequency point of view (Iacobucci and Noullez [11], Fidrmuc *et al*. [7], Poměnková *et al*. [19]). The theoretical background of the dynamic correlation and phase shift methods is nicely explained in the paper of Croux *et al*. [2] with a practical demonstration. One step ahead proposes time-frequency techniques combining both time approach and frequency approach. Thus, in one moment we are able to analyse behaviour of the time series from both perspectives. In the area of economic applications, a wavelet analysis is very popular (Ftiti [9], Berdiev *et al*. [1], Fidrmuc *et al*. [8]). The extension of co-movement measure into time-frequency domain in the form of cross-spectrum, co-spectrum, and cohesion propose more precise evaluation of co-movement to capture temporal character of co-movement (Maršálek *et al*. [18], Klejmová *et al*. [13], Kučerová and Poměnková [15]).

We focus on the credit standards for approving loans to enterprises and households of the Euro Area banks as an indicator of the supply of credit. Is there a relationship between the change of credit standards and the loan supply? A negative relationship between credit standards and new term loans to enterprises and households is expected. Therefore, the aim of the paper is to assess the co-movement between bank loans and credit standards in Euro Area via moving and dynamic correlation and also from the wavelet co-spectrum perspectives. We use quarterly data over the period 2003/Q1-2016/Q1.

2 Data and indicators

Data about the level of new loans (flows) provided to enterprises and households by MFIs (central banks, credit institutions and money markets funds) are drawn from the ECB Statistical Data Warehouse online database, item MFI aggregated balance sheet (ECB [4]). We also use data from the ECB Bank Lending Survey (BLS) (ECB [5]). The main aim of this survey is to enhance the understanding of bank lending behaviour in the euro area. As a measure of lending standards, we use a measure defined by ECB [7] – the diffusion index; it is based on the quarterly responses to questions concerning lending standards for the past three months (i.e. information from the supply side) and is measured as the weighted difference between the share of banks reporting that lending standards have been tightened and the share of banks reporting that they have been eased (banks who have answered "considerably" are given score 1 and banks that answered "somewhat" are given score 0.5). Positive values of the index indicate that a larger proportion of banks have tightened lending standards and vice versa. The representations of time series used in the analysis are depicted in Figure 1. The analysis is performed particularly for Euro Area countries (changing composition).

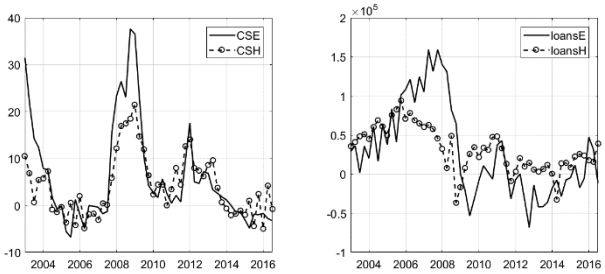


Figure 1 Time series representation, Source: ECB [4][5]

3 Methods

The relation between two time series can be measured via several approaches (Croux *et al*. [2]). The widely use is analysis in the time domain by correlation coefficient in the classic or the moving form. Because correlation method is well known, we skip its description here. In the frequency domain, we can use the dynamic correlation, the coherence or the cohesion. The dynamic correlation (Croux et al. [2])

$$\rho_{yz}(\omega_1,\omega_2)=\frac{\int_{\omega_1}^{\omega_2} C_{yz}(\omega)d\omega}{\sqrt{\int_{\omega_1}^{\omega_2} S_z(\omega)d\omega\int_{\omega_1}^{\omega_2} S_y(\omega)d\omega}}, \tag{1}$$

measure similarity of the frequency components of two time series *y* and *z*. Here (eq. (1)) *Cyz* is the real part of the cross-spectrum (i.e. co-spectrum) and *Sy*, *Sz* are the individual spectra of corresponding time series for frequencies ω. Integrating the eq. (1) in the frequency band from ω_1 to ω_2 evaluates the common behaviour of two time series in the given band of frequencies. If the bands are ω_1=0, ω_2 =π, the dynamic correlation coefficient corresponds to the classical correlation coefficient (Croux et al. [2], Kučerová and Poměnková[15]). Similar

analysis could be done in the time-scale domain via wavelet analysis and wavelet co-spectrum. The continuous wavelet transform (CWT) of the time series *y(t)* can be defined as (Jiang and Mahadevan [12])

$$S_{CWT}(a,b) = \int_{-\infty}^{\infty} y(t)\frac{1}{\sqrt{b}}\psi\left(\frac{t-a}{b}\right)dt, \quad b>0, a \in R, \tag{1}$$

where *a* is the time position, *b* is the parameter of dilatation (scale) of the mother wavelet $\psi(\cdot)$. To satisfy assumptions for the time-scale analysis, waves must be compact in time and frequency representation as well. Evaluation of the similarity between two time series *y*, *z* in the time-scale domain for their time-scale transform $S_{CWT,y}(a,b)$ and $S_{CWT,z}(a,b)$ can be calculated according to formula defined in Jiang and Mahadevan [12]

$$S_{12} = S_{CWT,1}(a,b)S_{CWT,2}(a,b) \tag{3}$$

If the time series are identical, then the co-spectrum is equal to one. For decreasing level of similarity the value of the co-spectrum approaches to zero.

4 Results

The analysis of the relationship between the credit standards imposed by banks on long-term loans provided to households (denoted as *CSH*) and to enterprises (denoted as *CSE*) and the volume of new loans (denoted as *loansE* and *loansH*) follows several steps, i.e. correlation analysis in classic and dynamic form and wavelet cospectrum. We also invetsigate lag 0 and lag 1 for new loans time series in order to identify whether there is a stronger relationship between credit standards in one quarter and new loans provided to households and to enterprises in the next quarter (lag 0) or in the period with the delay of two quarters (lag 1).[3]

	Lag 0				**Lag 1**			
	CSE	**loansE**	**CSH**	**loansH**	**CSE**	**loansE**	**CSH**	**loansH**
CSE	1	0,059	0,840***	-0,374***	1	0,211	0,851***	-0,276***
loansE	0,059	1	-0,113	0,568***	0,211	1	0,061	0,574***
CSH	0,840***	-0,113	1	-0,480***	0,851***	0,061	1	-0,406***
loansH	-0,374***	0,568***	-0,480***	1	-0,276***	0,574***	-0,406***	1

Table 1 Classic correlation coefficients

Note: statistically significant at: *** 1%, ** 5%, * 10%

In the first step, we perform a classical correlation analysis for the Q1/2003-Q1/2016 both with lag 0 and 1. The results are presented in Table 1. It is apparent that there is the significant negative correlation between the credit standards of banks and loans provided by banks to households in the analysed period, i.e. tighter credit standards are related to lower level of new loans provided. In case of loans to enterprises, the correlation is negative and significant only in lag 1.

In the next step, we calculate moving correlation illustrating the movement of correlation coefficient in the time (Fig. 2). It is clear that the moving correlation coefficient between credit standards and new loans to households without lags was negative (lower than –0.6) for most of the period, even after the financial crisis of 2007. However, it started to decrease in absolute value after the period Q2/2010 and has fluctuated between 0 and –0.2 since then. In case of enterprises without lag, the movement of this coefficient was more volatile with rise in the period from Q3/2003-Q3/2006 to Q1/2004-Q2/2007, then fall in the period from Q2/2004-Q2/2007 to Q1/2005-Q1/2008, then rise in the period from Q2/2005-Q2/2008 to Q4/2007-Q4/2010, then fall in the period from Q1/2008-Q1/2011 to Q2/2009-Q2/2012 and then was stable with another drop from Q1/2012-Q1/2015. In case of lag 1, the results concerning the trend are similar, i.e. relatively stable and negative correlation between credit standards and new loans to households with a change for the period after Q4/2008-Q4/2011 and volatile

[3] Credit standards are measured as backward looking changes of the bank behaviour concerning credit standards over the past three months. Therefore, we use the term "lag 0" in case of one quarter delay and "lag 1" in case of two quarters delay. The reason is that the credit standards are reported in the period following the period with the surveyed change of credit standards (i.e. data concerning the changes of credit standards in Q2 are reported in Q3). In our paper, we measure the relationship between these credit standards reported in the following period (i.e. in Q3) and new loans provided in the same period in case of lag 0 (i.e. in Q3) or in the next period in case of lag 1 (i.e. in Q4). From the statistical point of view, there is no lag or a lag of one quarter. However, from the economic point of view, there is a lag of one quarter or two quarters.

positive or negative correlations between credit standards and new loans to enterprises when negative correlations are measured in the post-crisis period (from Q4/2008-Q4/2011).

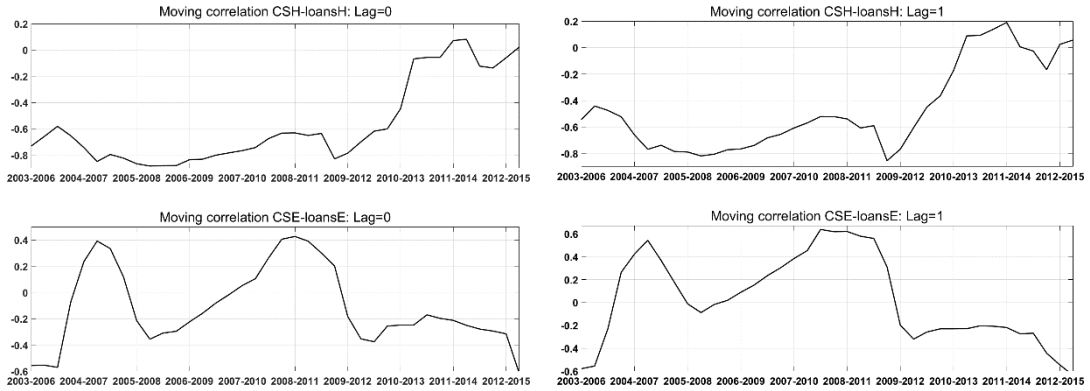


Figure 2 Moving correlation

However, both classical and moving correlation gives us overview about the mutual linear (positive or negative) dependence between indicators. Via the moving correlation we can distinguish the sub-periods of the correlation movements which can help us identify moments of changes. Even though both approaches give reasonable results and confirm negative dependence between lending standards and loans provided by banks in the analysed period, we proceed in dynamic correlation (correlation in frequency domain) to obtain detailed picture about dependence structure with respect to the movement in the time series structure.

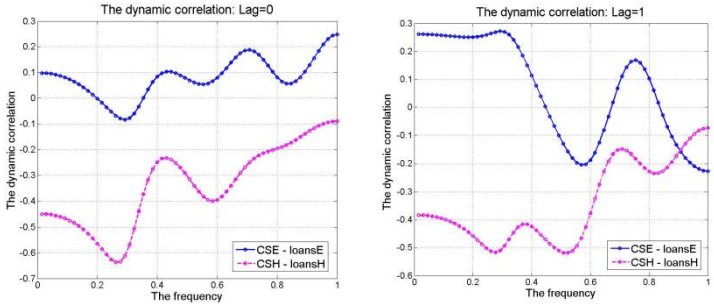


Figure 3 Dynamic correlation

The results of the dynamic correlation between *CSE* and *loansE*, *CSH* and *loansH* for lag 0 and 1are in Fig. 3. Focusing on results for lag 0, both dynamic curves have similar shape but in the different level. The dependence between *CSH* and *loansH* is moreover significant across all frequencies and thus dynamic correlation confirms the results given by classic correlation. The dependence between *CSE* and *loansE* is, unfortunately, insignificant. Investigating one lag order, we see the different shape of curves. We can distinguish several sub-periods. In the first (range between 0-0.35), there is significant both positive (*CSE* and *loansE*) and negative (*CSH* and *loansH*) dependence. Such sub-period covers business cycle frequencies, i.e. nested cycles from 32 quarters (time-trend and long cycles) as well as short cycles of the 6 quarter length. After that both dynamic curves have opposite tendency in the second sub-period (range between range between 0.35-0.4 and 0.45-0.6) covering short fluctuations (short cycles of the length from 6 to 3 quarters). Finally, both curves indicate insignificant dependencies in the range 0.62-1 which denote rapid changes. Hence, we can conclude that there is a stronger negative dependence between credit standards and new loans to households and this behaviour of banks includes the long-term to particularly short-term movements while in case of enterprises the dependence is mostly positive and follows mostly business cycle frequencies.

With respect to the possible structural breaks in time series, we proceed with the last step which is a co-spectral analysis. The co-spectral analysis allows us to assess the dependence and to identify areas of the comovement between time series from time as well as frequency perspective in one moment. From the family of mother wavelet functions, we use the Morlet wavelet. Results can be seen in Fig. 4. The co-spectrum is denoted on the *z*-axis for specific periods (*x*-axis; time) and periodicities (*y*-axis; periods in quarters). The figure shows a two-dimensional projection of three-dimensional charts. The intensity of different contours denotes the relative importance of different periodicities and time. In all figures (Fig. 4), the wavelet co-spectrum shows several areas of comovement bounded by the time and frequency. The lower is the frequency of inputs (i.e. the higher and closer to the value of 36 the figure on the *y*-axis) the longer is the cyclical component. Denote that the business cycles frequencies are defined between 6 quarters (rapid moving periodic component) and 8 years (slow moving periodic component).

For co-spectrum of *CSE-loansE* and *CSH-loansH* and lag 0, we can observe same areas, i.e. 2008-2010 and 2013-2015 covering frequencies till 30 quarters, 20 quarters respectively. In addition, for co-spectrum of *CSH-loansH* we can also see area covering higher frequencies, i.e. Q3/2008-Q2/2009 and 2011-2012 covering frequencies from 25 to 16 quarters and from 27 to 14 quarters respectively. In the case the co-spectrum of *CSE-loans E* and lag 1, we can observe areas 2006-2007, Q3/2008-Q2/2009, 2011-2012 covering frequencies till 30 quarters and 2013-2015 covering frequencies till 20 quarters. For the co-spectrum of *CSH-loansH* and lag 1, we can observe two similar areas as in previous case, i.e. 2008-2010 covering frequencies till 30 quarters and 2013-2015 covering frequencies till 20 quarters. In addition, we can identify area Q2/2008-Q2/2010 covering frequencies from 26 to 15 quarters and the area Q2/2011-Q2/2012 covering frequencies from 26 to 14 quarters.

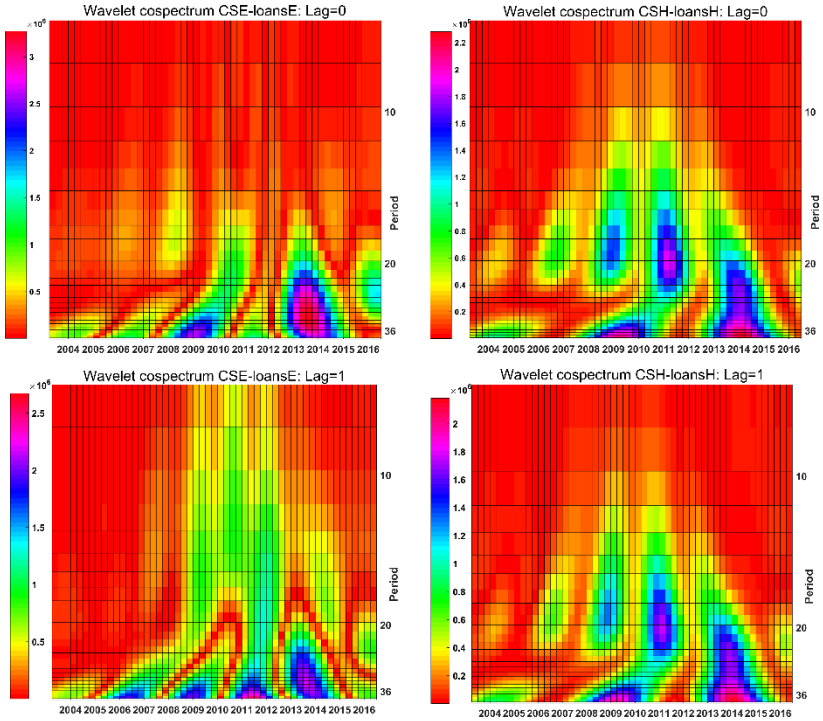


Figure 4 Wavelet co-spectrum

To sum it up, we can conclude that the wavelet analysis specify results given by the application of a dynamic correlation from the time perspectives. We can also conclude, that reactions of households reflect predominantly long-term in straight reaction. The quick changes and very short cyclical movements reaction was visible with respect to the lag investigating. In case of enterprises, the significant changes are mostly in the long-term cycles. In case of households, reaction on quick changes and very short cyclical movements were outlined and consequently confirmed by lag investigation. In both cases, we were able to specify the time range of such co-movements.

5 Conclusion

The paper was focused on the role of credit standards for approving loans to enterprises and households of the Euro Area banks as an indicator of the supply of credit. Therefore, we wanted to answer the question whether there is a relationship between the change of credit standards and the loan supply or not. The aim of the paper was to assess the co-movement between bank loans and credit standards in Euro Area via moving and dynamic correlation and also from the wavelet co-spectrum perspectives.

A negative relationship between credit standards and new term loans to enterprises and households was expected. In our analysis, we confirmed this relationship mainly in case of loans to households. However, this relationship was negative and stable only before the financial crisis. In case of enterprises, the hypothesis was not confirmed; the results for classical correlation were not significant and for moving correlation were quite volatile. In the period around financial crisis, the relationship was even positive; banks strongly tightened their credit standards in reaction to the crisis in 2007 and partly in 2008 but then they eased these standards strongly in 2009 and 2010. However, the volume of new loans provided to enterprises dropped dramatically in 2008-2009.

The dynamic correlation analysis confirms the expected stronger negative dependence between credit standards and new loans to households, i.e. this behaviour of banks includes the long-term to particularly short-term movements while in case of enterprises this behaviour follows mostly business cycle frequencies and do not reflect

very short reactions. The wavelet analysis helped with specification of the results given by the application of the dynamic correlation. We identify several areas limited by time and frequency leading to several conclusions. We found that credit standards and new loans to enterprises, households respectively, from the perspectives of the long-term co-movement was the strongest between 2008-2010 and 2013-2015. In case of credit standards and new loans to households, there are additional areas of short-term co-movement between 2006-2007 and 2011-2012. We can also conclude that reactions of households reflect quick changes and very short cyclical movements.

Acknowledgements

The research described in the paper was supported by the Czech Science Foundation via grant n. 17-24309S and by the Czech Ministry of Education in the frame of National Sustainability Program under grant LO1401. For research, the infrastructure of the SIX Center was used.

References

[1] Berdiev, A. N., Chang, Ch-P.: Business cycle synchronization in Asia-Pacific: New evidence from wavelet analysis. *Journal of Asia Economics* **37** (2015), 20-33.
[2] Croux, Ch., Forni, M., Reichlin, L.: A Measure of Comovement for Economic Variables: Theory And Empirics. *The Review of Economics and Statistics* **83** (2001), 232-241.
[3] Engle, R. F., Granger, C. W.: Cointegration and error correction: representation, estimation and testing. *Econometrica* **55** (1987), 251–276.
[4] European Central Bank. Statistical Data Warehouse: MFI Aggregated Balance Sheet , (2016a). Retrieved March 27, 2017, from http://sdw.ecb.europa.eu/browse.do?node=9691115.
[5] European Central Bank: Statistical Data Warehouse: Bank Lending Survey, (2016b). Retrieved March 27, 2017, from https://sdw.ecb.europa.eu/browse.do?node=9484572.
[6] Feldkircher, M., Korhonen, I. The rise of China and its implications for the global economy:evidence from a global vector autopregressive model. *Pacific Economic Review* **19** (2014):1
[7] Fidrmuc, J., Ikeda, T., Iwatsubo, K.: International transmission of business cycles: Evidence from dynamic correlations. *Economic Letters* **114** (2012), 252-255.
[8] Fidrmuc, J., Korhonen, I., Poměnková, J.: Wavelet spectrum analysis of business cycles of China and G7 countries. *Applied Economic Letters* **21** (2014), 1309-1313.
[9] Ftiti, Z., Tiwari, A., Belanés, A.: Tests of Financial Market Contagion: Evolutionary Cospectral Analysis V. S. Wavelet Analysis. *Computational Economics* **46** (2014), 575-611.
[10] Hassan, G., Cooray, A., Holmes, M. The effect of female and male health on economic growth: cross-country evidence within a production function Framework, *Empirical Economics* **52** (2017), 659-689
[11] Iacobucci A., Noullez, A.: A frequency selective filter for short-length time series. *Computational Economics* **25** (2005), 75–102.
[12] Jiang, W., Mahadevan, S.: Wavelet spectrum analysis approach to model validation of dynamic systems. *Mechanical Systems and Signal Processing* **25** (2011), 575–590.
[13] Klejmová, E., Poměnková, J., Blumenstein, J.: Combination of time-frequency representation for background noise suppresion, In: *31st European Conference on Modelling and Simulation ECMS 2017*. Forthcomming.
[14] Korhonen, I., Peresetsky, A. What influences stock market behavior in Russia and other emerging countries, *Emergin Market Finance & Trade* **52** (2016), 1210-1225.
[15] Kučerová, Z., Poměnková, J.: Financial and Trade Integration of Selected EU Regions: Dynamic Correlation and Wavelet Approach. *Ekonomický časopis* **63** (2015), 686-704.
[16] Lukmanova, E, Tondl, G.: Macroeconomic imbalances and business cycle synchronization. Why common economic governance is imperative for the Eurozone. *Economic Modelling 62* (2017), 130-144.
[17] Lu, F., Qiao, H., Wang, S., Lai, K. K., Li, Y.: Time-varying coefficient vector autoregressions model based on dynamic correlation with an application to crude oil and stock markets. *Environmental Research* **152** (2017), 351-359.
[18] Maršálek, R., Poměnková, J., Kapounek, S.: A Wavelet-Based Approach to Filter Out Symmetric Macroeconomic Shocks. *Computational Economics* **44** (2014), 477-488.
[19] Poměnková, J., Kapounek, S., Maršálek, R.: Variability of Dynamic Correlation – the Evidence of Sectoral Specialization in V4 Countries. *Prague Economic Papers* **3** (2014), 371-384.

Multidimensional Alpha Stable Distribution in Model Parameter Estimation Algorithms

Jaromír Kukal[1], Quang Van Tran[2]

Abstract.
Solving optimization problem with a multidimensional objective function is the traditional task of operational research. Both multimodality and nonsmoothness of the objective function are typical difficulties when searching for the global optimum. To improve the search effectiveness in non-adaptive algorithms as well as the ability to leap away from local optima, the so called Levy flights are often used. The Levy flights are random non-Gaussian step sizes following alpha stable distribution. However, this random variable is very difficult to be generated in the multivariatal case, the step size generation procedure has never been treated accordingly so far. To examine the effectiveness of the general usage of this possibility, the goal of this paper is to introduce the multivariate alpha stable random variable step size generation technique into several novel optimization algorithms. Our approach then is used to estimate parameters of generalized hyperbolic distribution with the use of maximum likelihood estimation method for returns of Czech stock market index PX from 2000 to 2017. As this task is almost unsolvable for traditional optimization methods, the results we obtained are quite promising.

Keywords: Optimization, Alpha stable distribution, Multidimensional search, Heuristic algorithm, Parameter estimation.

JEL classification: C44

1 Introduction

Traditional optimization methods are ineffective when solving a problem with a non-convex, multimodal and discontinuous objective function. In this case in order to reach the required global solution, global optimization algorithms are applied. For these algorithms their performance heavily depends on the proper mutation creation as it can improve the searching effectiveness as well as the ability to leap away from local optima. For this purpose, Levy flight mutation technique is often used. As the Levy flights are a non-Gaussian random variable whose generation is very difficult in the multivariatal case, the question how the generation procedure can affect the searching effectiveness has not been appropriately investigated so far. As an attempt to give an answer to this question, this paper will introduce a general multivariate alpha stable random variable step size generation technique into several advanced optimization algorithms, namely Random Descent improved with Lévy flight algorithm (RDLF), Cuckoo Search (CS) and Modified Cuckoo Search algorithms (MCS). The effectiveness of our approach then will be verified on the estimation parameters of generalized hyperbolic distribution by the maximum likelihood estimation method. The likelihood function for this as objective function is well-known for its complicated shape, hence in this experiment we will try to evaluate not only the effect of searching parameters, but also of stable distribution on reaching the global optimum. The parameters will be identified for returns of daily Czech stock market index PX from 2000 to 2017.

2 Standard Multivariate α-Stable Distribution Simulation

The class of stable distributions plays role in modern statistical theories. The random vectors of this class are frequently used in many applications. A random vector $\mathbf{X} = (X_1, \ldots, X_d)^{\mathrm{T}}$ is said to be a stable random vector in $\mathbb{R}^d$ if for any positive numbers $A_1, \ldots, A_n$ there is a positive number B and a vector $\mathbf{c}$ in $\mathbb{R}^d$ such that

$$A_1\mathbf{X}^{(1)} + \ldots + A_n\mathbf{X}^{(n)} \stackrel{d}{=} B\mathbf{X} + \mathbf{c}, \tag{1}$$

where $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ are independent and from the same distribution as $\mathbf{X}$ (for more detailed information, see [5]). For the sake of notation simplicity without the loss of generality, here we will deal only with the case of standard multivariate alpha stable distribution. Further, due to the complexity of covariance structures we also focus only

[1]Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Software Engineering, 115 19 Břehová 7, Prague, Czech Republic , jaromir.kukal@fjfi.cvut.cz
[2]Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Department of Software Engineering, 115 19 Břehová 7, Prague, Czech Republic, tranvqua@fjfi.cvut.cz

on the most tractable case when the iso-density plot of joint density forms a circle. The characteristic function of an isotropic standard stable variable (also called radially symmetric stable random variable) is as follows

$$\psi(\mathbf{u}) = \mathrm{E}\exp(i\mathbf{u}^{\mathrm{T}}\mathbf{X}) = \exp(-\mid \mathbf{u} \mid^{\alpha}), \tag{2}$$

where

$$\mathbf{X} = A^{\frac{1}{2}}\mathbf{G} \tag{3}$$

where $\mathbf{G} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ and $A \sim \mathrm{S}(\alpha/2, 1, 2\cos(\alpha\pi/4)^{2/\alpha}, 0)$, for $0 < \alpha < 2$ and S is α stable distribution [3]. The process of standardized generation will be denoted as $\mathbf{X} \sim \mathrm{L}(\alpha, d)$. It is clear that in order to generate a multivariate stable variable, one must generate univariate stable variable. A good algorithm for generating univariate standard stable random variable $Y \sim \mathrm{S}(\alpha, \beta, 1, 0)$ is proposed by Chambers et al [2] as follows

- Generate $U \sim \mathrm{U}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and $W \sim \mathrm{Ex}(1)$
- Compute for $\alpha \neq 1$

$$Y = (1+\zeta^2)^{\frac{1}{2\alpha}} \frac{\sin(\alpha(U+\xi))}{(\cos(U))^{\frac{1}{\alpha}}} \left(\frac{\cos(U-\alpha(U+\xi))}{W}\right)^{\frac{1-\alpha}{\alpha}} \tag{4}$$

- for $\alpha = 1$

$$Y = \frac{1}{\xi}\left[\left(\frac{\pi}{2}+\beta U\right)\tan(U) - \beta\left(\frac{\pi W\cos(U)}{\pi+2\beta U}\right)\right], \tag{5}$$

where $\zeta = -\beta\tan\frac{\alpha\pi}{2}$ and $\xi = \begin{cases}\alpha^{-1}\arctan(-\zeta) & \text{if } \alpha \neq 1\\ \pi/2 & \text{otherwise}\end{cases}$.

3 Levy Flights in Optimization Algorithms

A Lévy flight is a random variable which has a distribution with heavy-tails. The term was first used by Mandelbrot to describe the distribution of step sizes that follow a Cauchy distribution. Both Cauchy distribution and normal distribution are special cases of alpha stable distribution. However, the Cauchy distribution has very strong tails as the probability of extreme events decreases polynomially compared to the normal distribution when it falls exponentially. This property has recently widely exploited in optimization algorithms to improve their performance. Lévy flights are used to generate new mutations. Since the step sizes with Levy flight tend to be longer, they allow a better randomization in algorithms as well as more efficient exploration of the searching domain. Longer step sizes also prevent the false convergence to local optima and increase the probability of reaching global optimum. The Lévy flight is directly applicable in optimization tasks related to nonlinear regression, complex physical and chemical equilibria investigations and operation research tasks.

For a given domain $\mathcal{D} = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}$, of dimension $d \in \mathbb{R}$ where $\mathbf{a}$ and $\mathbf{b}$ are the boundaries we will minimize the objective function $\mathrm{f} : \mathcal{D} \to \mathbb{R}$. Formally, the Lévy flight mutation works as follows. Using mutation temperature $T_{\mathrm{mut}} > 0$ and perturbation function $\mathrm{P} : \mathbb{R}^d \to \mathcal{D}$, the Lévy flight mutation generates $\mathbf{x}_{\mathrm{new}} \in \mathcal{D}$ from the original $\mathbf{x} \in \mathcal{D}$ as $\mathbf{x}_{\mathrm{new}} = \mathrm{P}(\mathbf{x} + T_{\mathrm{mut}}\mathbf{s})$ where $\mathbf{s} \sim \mathrm{L}(\alpha, d)$ and the perturbation P is implemented as component–wise boundary mirroring whenever the component x_j is outside $[a_j, b_j]$. The selection of α, T_{mut} is the subject of our investigation in next sections.

3.1 Random Descent with Lévy Flight Algorithm

Traditional Random Search algorithm (RD) is based on short range mutation with constrained variance which is frequently realized using multidimensional uniform or Gaussian distribution around $\mathbf{x}$. This first choice algorithm is able to find only local minima of objective function. There is a nice opportunity for Lévy flight mutation to offspring this disadvantage of RD and increase the probability of global optimum finding. The final form of novel heuristics with Lévy flight mutation is explained as Algorithm 1.

3.2 Cuckoo Search

Cuckoo Search (CS) optimization algorithm is inspired by cuckoo's behavior in nature. For a given cuckoos' population $N \in \mathbb{N}$ a bird nest $\mathbf{x}_k \in \mathcal{D}$ for $k = 1, \ldots, N$, whose quality is evaluated by the objective function as $f_k = \mathrm{f}(\mathbf{x}_k)$. A traditional algorithm has five basic operations: initialization, local search, sorting, abandoning, and termination. At the initialization stage N occupied nests are randomly generated with from an uniform distribution and their subsequent evaluation as

$$\mathbf{x}_k \sim \mathrm{U}(\mathcal{D}) \text{ with } f_k = \mathrm{f}(\mathbf{x}_k). \tag{6}$$

At the local search stage two nests are randomly selected and a cuckoo flies from the i-th nest to a new position with the Lévy flight mutation described in the previous section above as

$$\mathbf{y} = \mathrm{P}(\mathbf{x}_i + T_{\mathrm{mut}}\mathbf{s}) \in \mathcal{D}. \tag{7}$$

Algorithm 1 Random Descent with Lévy Flight

function RDLF
  $\mathbf{x} \sim \mathrm{U}(\mathcal{D}), f = \mathrm{f}(\mathbf{x})$
  loop
    $\mathbf{s} \sim \mathrm{L}(\alpha, d)$
    $\mathbf{x}_{\mathrm{new}} = \mathrm{P}(\mathbf{x} + T_{\mathrm{mut}}\mathbf{s})$
    $f_{\mathrm{new}} = \mathrm{f}(\mathbf{x}_{\mathrm{new}})$
    if $f_{\mathrm{new}} < f$ **then**
      $\mathbf{x} = \mathbf{x}_{\mathrm{new}}$
      $f = f_{\mathrm{new}}$

The competition between $\mathbf{y}$ and the j-th nest is determined by the rule

$$\mathrm{f}(\mathbf{y}) < f_k \Rightarrow \mathbf{x}_j = \mathbf{y}, f_j = \mathrm{f}(\mathbf{y}). \tag{8}$$

At the sorting stage the population of occupied nests is sorted in such a way so that for any pair of nests they must satisfy

$$k < l \Rightarrow f_k \leq f_l. \tag{9}$$

At the abandoning stage, M nests with highest objective function value are removed and replaced by new M nests using constrained Lévy fight mutation without competition. Their new positions and the corresponding qualities are

$$\mathbf{x}_k^{\mathrm{new}} = \mathrm{P}(\mathbf{x}_k + T_{\mathrm{mut}}\mathbf{s}) \in \mathcal{D}, f_k^{\mathrm{new}} = \mathrm{f}(\mathbf{x}_k^{\mathrm{new}}), \text{ for } k = N - M + 1, \ldots, N. \tag{10}$$

At the final termination stage a Cuckoo Search algorithm can be stopped by touching the optimal solution vicinity, population stagnation or by reaching the preset maximal number of iterations. According to [7], Cuckoo Search algorithm can perform better on testing examples than differential evolution or particle swarm optimization algorithm in terms of time complexity and reliability. Cuckoo Search algorithm has been later modified into a more sophisticated version called Modified Cuckoo Search (MCS, see [6]). In this modification the initialization and searching steps have been improved to speed up the searching procedure.

4 Minimization of Non-convex Negative Log-likelihood Function

Maximum Likelihood Estimation technique (MLE) is one of most often used methods to estimate the parameters of a model for given observations. For a given dataset and model specification this method tries to find the parameter values that maximize the likelihood function value. This means that it maximizes the matching of the selected specification with the observed data. Mathematically, given a dataset $\mathbf{x}_k \in \mathbb{R}^n, k = 1, \ldots, m$ and the model density $\mathrm{f}(\mathbf{x}; \mathbf{w})$, where $\mathbf{w} \in \mathcal{D} \subset \mathbb{R}^n$ is a vector of parameters, the log-likelihood function is

$$\Phi(\mathbf{w}) = \ln L(\mathbf{w}) = \sum_{k=1}^{m} \ln \mathrm{f}(\mathbf{x}_k; \mathbf{w}). \tag{11}$$

The maximum likelihood estimate on $\mathcal{D}$ is

$$\mathbf{w}_{\mathrm{opt}} \in \underset{\mathbf{w} \in \mathcal{D}}{\operatorname{argmax}}\, \Phi(\mathbf{w}). \tag{12}$$

In practice, instead of maximizing the log-likelihood function, for convenience we minimize the Negative Log-Likelihood function (NLL).

To find the optimal solution to this problem, any traditional iteration minimization algorithm can be used. However, the negative log-likelihood function might be non-convex and non-smooth, which makes them get stuck to a local minimum and disallows us to reach the optimal global solution. This gives rise to various alternative optimization procedures and some of the most powerful ones have been described in the previous sections.

5 A Case Study: Generalized Hyperbolic Distribution of Returns

It is well-known that returns of financial assets are not normally distributed and they tend to have thicker tails and higher kurtosis. This feature can be captured by the Generalized Hyperbolic Distribution (GHD), which was first introduced by Barndorff-Nielsen et al. [1]. This distribution is characterized by five parameters $\theta = (\lambda, \alpha, \beta, \delta, \mu)$, where $\alpha, \delta \in \mathbb{R}^+$, $\lambda, \mu \in \mathbb{R}$, $\mid \beta \mid < \alpha$ and its probability density function is

$$f_{\mathrm{GH}}(x; \theta) = \kappa[\delta^2 + (x - \mu)^2]^{\frac{1}{2}\left(\lambda - \frac{1}{2}\right)} \mathrm{K}_{\lambda - \frac{1}{2}}\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right) \exp(\beta(x - \mu)), \tag{13}$$

where $\kappa = \dfrac{(\alpha^2-\beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi}\alpha^{\lambda-\frac{1}{2}}\delta^{\lambda}\mathrm{K}_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}$, and K_{λ} is the modified Bessel function of the third kind of order λ.
All moments of a random variable from a generalized hyperbolic distribution exist and the first two of them are

$$\mathrm{E}(X) = \mu + \frac{\beta\delta^2}{\zeta}\frac{\mathrm{K}_{\lambda+1}(\zeta)}{\mathrm{K}_{\lambda}(\zeta)} \tag{14}$$

$$\mathrm{Var}(X) = \delta^2\left\{\frac{\mathrm{K}_{\lambda+1}(\zeta)}{\mathrm{K}_{\lambda}(\zeta)} + \left(\frac{\beta\delta}{\zeta}\right)^2\left[\frac{\mathrm{K}_{\lambda+2}(\zeta)}{\mathrm{K}_{\lambda}(\zeta)} - \left(\frac{\mathrm{K}_{\lambda+1}(\zeta)}{\mathrm{K}_{\lambda}(\zeta)}\right)^2\right]\right\} \tag{15}$$

where $\zeta = \delta\sqrt{\alpha^2-\beta^2}$. The fat tail property of the GHD originates from the fact

$$\mathrm{P}(X<x) = \mathrm{O}(|x|^{\lambda-1}\exp((\alpha+\beta)x)) \text{ as } x \to \pm\infty. \tag{16}$$

The interpretation of each parameter of GHD is as follows. The mean and variance of the GDH are already shown in (14) and (15). For higher moments, after transformations $\tau = \dfrac{\beta}{\gamma}$, $\zeta = \delta\gamma$, $\chi = \dfrac{\beta}{\alpha\sqrt{1+\zeta}}$, $\xi = \dfrac{1}{\sqrt{1+\zeta}}$, where $\gamma = \sqrt{\alpha^2+\beta^2}$, then according to [1] the parameters χ and ξ are considered to be the natural measures for skewness and kurtosis.

We use our novel approaches described in the previous sections to estimate parameters of generalized hyperbolic distribution for returns of Prague stock market index PX. Data is a series of its daily close values from January 2000 to April 2017. The original data is transformed into a series of logarithmic returns. Some of their descriptive statistics are shown in Table 1. Here the most standing-out number is the value of returns kurtosis which far more exceeds the value for the normal distribution.

Series	Mean	Median	Maximum	Minimum	Std. dev	Skewness	Kurtosis
index	982.29	970.82	1936.10	320.10	369.91	0.34	2.67
returns	1.62×10^{-4}	0.00055	0.1236	-0.1619	0.0140	-0.47	15.52

Table 1 Descriptive statistics of series PX and its returns

Parameters of GHD distribution are estimated by maximum likelihood estimation method. The density function defined in (13) is used to set up the the negative log-likelihood function. The objective function is known for its non-convex course [4] and hence it asks for the use of progressive optimization methods. Our optimization algorithms are applied to estimate the parameters. The estimation results including the standard errors of the estimates are displayed in Table 2. As seen, all estimated parameters are significant from zero. Parameter λ also significantly differs from values of the hyperbolic and the inverse gaussian distributions. In Figure 1 we compare the matching of GHD distribution to data with that of the corresponding normal distribution where their left tails are zoomed out in the right panel. The GHD distribution clearly matches the data better than the normal distribution.

Parameter	λ	α	β	δ	μ
Value	-1.7591	7.0825	-5.0329	0.0173	0.0011
S.E.	0.0384	0.0533	0.7574	0.0012	2.76×10^{-4}

Table 2 GHD Parameters Estimation Results

Regarding the impact of the shape of stable distribution to mutation generation and optimization parameters on results, we proceed as follows. Values of α of stable distribution is set at: 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, and 1.95. Values of T_{mut} are: 0.001, 0.002, 0.005, 0.02, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, and 10. Further, f^* is set at 99.9% of the best reachable value of objective function from 1000 independent runs of RDLF which is a good approximation of the global minimum. The maximum number of evaluations per run was 2000. The time complexity of searching was studied using 100 independent runs of the same length to obtain time complexity as mean value of function evaluation number for given T_{mut}, α. As parameters of GHD distribution have different ranges, in order to make the search in the searching domain more efficient, we standardize the searching range in the following way: $\lambda = \mathrm{atanh}(\xi_1)$, $\alpha = -\log(1-\xi_2)$, $\beta = \alpha\xi_3$, $\delta = -\log(1-\xi_4)$ and $\mu = \mathrm{atanh}(\xi_5)$, where $\xi_1, \xi_5 \in [-1+\epsilon, 1-\epsilon]$, $\xi_2, \xi_3, \xi_4 \in [\epsilon, 1-\epsilon]$, $\epsilon = 10^{-12}$. Except of T_{mut}, α, RDLF has no additional parameter. The remaining parameters of CS were set to $N = 15$, $M = 4$ as recommended in [7]. The MCS was used for $N = 20$ as population size, $NG = 15$ as the abandoned nest number, and $NL = 5$ as the elite nest number suggested by [6]. The effect of α and T_{mut} on the expected number of evaluation in selected algorithms are displayed in Tables 3, 4 and 5.

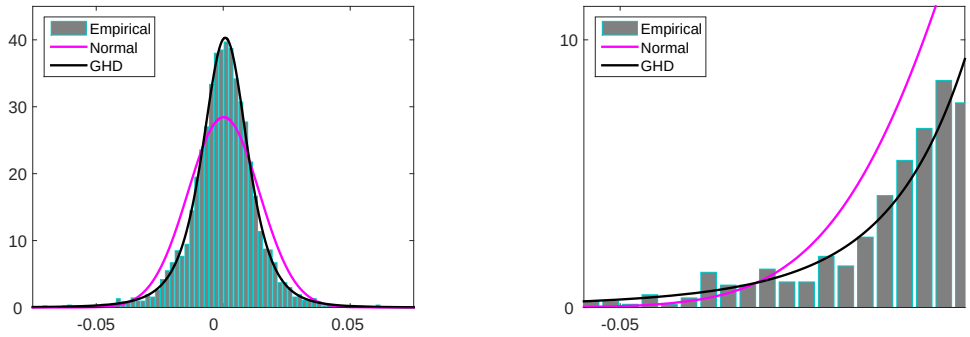


Figure 1 The estimated PDFs of analyzed distributions of EUR/CZK returns

T_{mut}	α							
	0.5	0.6	0.7	0.8	0.9	1	1.5	1.95
0.001	234	285	390	466	549	697	1118	1098
0.002	182	223	265	322	388	497	852	982
0.005	124	145	165	178	201	276	397	416
0.01	114	112	116	128	143	154	267	269
0.02	100	95	93	88	89	108	158	205
0.05	73	75	73	76	78	72	79	100
0.1	85	68	74	70	71	69	66	60
0.2	95	101	106	79	85	74	104	95
0.5	128	122	154	163	163	188	213	246
1	177	182	226	271	306	318	392	417
2	268	321	405	467	469	526	537	495
5	472	545	466	493	533	621	483	496
10	516	444	516	552	523	582	579	489

Table 3 Time complexity of Random Descent with Lévy Flight

T_{mut}	α							
	0.5	0.6	0.7	0.8	0.9	1	1.5	1.95
0.001	787	877	923	880	1016	1023	846	497
0.002	704	803	785	866	1018	941	862	945
0.005	595	590	594	812	818	818	1133	1045
0.01	477	551	589	558	631	625	817	984
0.02	455	474	473	453	478	554	694	872
0.05	405	394	324	404	408	388	454	496
0.1	432	356	323	373	340	388	329	375
0.2	430	390	399	307	368	408	347	367
0.5	374	414	413	396	420	407	450	422
1	428	479	498	447	589	580	569	436
2	579	508	422	538	516	463	498	544
5	478	407	490	569	547	567	541	521
10	511	509	508	542	523	517	549	475

Table 4 Time complexity of Cuckoo Search

In the case of RDLF heuristics (Table 3), the best results were obtained for $T_{\text{mut}} \in [0.05, 0.1]$ and any $\alpha \leq 1$. This indicates that if the temperature is well chosen, the role of Lévy flight shape is less important. The method is reliable for any $T_{\text{mut}} \in [0.005, 5]$ and $\alpha \leq 1$. When the temperature is not optimal, a lower α up to 0.8 can help.

With CS (Table 4), for $T_{\mathrm{mut}} \geq 0.01$, any value of α is acceptable and the threshold is most comfortably achieved with $T_{\mathrm{mut}} \in [0.05, 1]$. The best obtained time complexity value is four times slower than with RDLF. For MCS, with $\alpha = 0.5$, any $T_{\mathrm{mut}} \geq 0.002$ is suitable and on the contrary, for $T_{\mathrm{mut}} \geq 0.01$, any value of α can lead to the successful optimization completion. The best time complexity value in this case is three times slower than that of RDLF (Table 5).

T_{mut}	α							
	0.5	0.6	0.7	0.8	0.9	1	1.5	1.95
0.001	281	348	268	294	343	342	375	405
0.002	304	260	269	287	291	273	307	376
0.005	246	252	232	232	265	230	238	232
0.01	240	233	224	240	208	208	222	205
0.02	238	224	206	213	250	219	214	195
0.05	219	215	217	232	236	205	194	173
0.1	211	202	246	210	234	206	228	219
0.2	195	232	260	224	201	203	213	238
0.5	218	221	214	188	227	221	198	230
1	208	201	198	217	222	237	214	211
2	219	198	214	198	173	233	215	201
5	202	217	205	210	227	218	209	171
10	192	241	224	193	250	179	208	220

Table 5 Time complexity of Modified Cuckoo Search

6 Conclusion

Alternative optimization algorithms have become an indispensable tool for finding a solution to a global optimization problem when traditional methods fail to succeed. For these algorithms the creation of appropriate mutations is an important operation and it can help to multiply increase their searching effectiveness. In this experiment, we have investigated the possible use of alpha stable distribution to randomize the step size in mutation creation for three advanced non-adaptive searching heuristics: RDLF, CS and MCS. This approach is then verified on the task of estimating parameters of generalized hyperbolic distribution for daily returns of stock market index PX, whose log-likelihood function is not well-behaved. The obtained results show that in addition to the traditional role of temperature, the shape of alpha stable distribution can substantially increase the searching capacity of these algorithms in many cases.

Acknowledgements

This paper is supported by grant SGS17/196/OKH4/3T/14.

References

[1] Barndorff-Nielsen, O. E., Bloesild, P., Jensen, J. L., and Sørensen, M.: The fascination of sand. In: *A Celebration of Statistics*, (Atkinson, A. C., Fienberg, S. E., ed.) 1985, Springer Verlag, New York, pp. 57—87.

[2] Chambers, J. M., Mallows, C. L., and Stuck, B. W.: A Method for Simulating Stable Random Variables. *Journal of the American Statistical Association* 71 (1976), 340–344.

[3] Nolan, J. P.: Multivariate elliptically contoured stable distributions: theory and estimation. *Computational statistics* **28**/5 (2013), 2067–2089.

[4] Prause, K.: *The generalized hyperbolic model: Estimation, financial derivatives and risk measures.* University of Freiburg, Doctoral Thesis, 1999.

[5] Samorodnitsky, G., Taqqu, M.: *Stable non-Gaussian random processes*. Chapman and Hall, New York, 1994.

[6] Walton, S., Hassan, O., Morgan, K., and Brown, M.,R.: Modified Cuckoo Search: A New Gradient Free Optimisation Algorithm. *Chaos, Solitons & Fractals* 44 (2011), 710-718.

[7] Yang, X.,S., and Deb, S.: Engineering Optimisation by Cuckoo Search, *International Journal of Mathematical Modelling and Numerical Optimisation* 1 (2010), 330-343.

Optimization Model for the Employees' Shifts Schedule

Martina Kuncová[1], Veronika Picková[2]

Abstract. Optimization models can be useful and helpful in various real-life situations where any other methods or techniques are hard to apply to find the solution. The staff or shift scheduling problem belongs to these situations as the problem is complex thanks to many different constraints. This problem can be modeled on the bases of the managers' requirements that are influenced by customers' demand, number of potential employees working in shifts, employees' qualification, number and lengths of shifts, days off or various wishes/limitations of employees. This article describes the solution of the real employee shift work scheduling problem via optimization model. The model is based on data that are given by the enterprise manager (especially the employees' work attendance for the given week). The integer linear programming with the assignment optimization model are mixed to find the optimal solution respecting given constraints. The model was solved by LINGO solver with MS Excel usage for the data settings but also the MS Excel macro was prepared for the managers' needs.

Keywords: Optimization model, shift work, scheduling, rosters.

JEL Classification: C44, M12
AMS Classification: 90B35, 90C10

1 Introduction

Scheduling problems as a task to assign a person to a given position or given time or shift belong to the problems that can be solved via various technics of operational research. Sometimes only Gantt chart is enough for adding a person to an activity and to see the working hours for each person or to display the roster graphically [2]. But with rising number of persons and different shifts or shift lengths and with other conditions being met it is better to use an optimization model. Usually these models are based on linear programming (LP), integer (ILP) or mixed-integer programming (MILP), assignment problem, stochastic programming, non-linear programming or goal programming [1] but also other solution methods are possible such as demand modeling, metaheuristics [2] or constructive and improvement heuristics [7]. They are usually called shift scheduling problems, rostering problems [2], personnel scheduling problems [7] or timetabling problems (that are slightly different and connected with the time slots more than shifts) [6]. According to [4] two main variations of the problem can be solved: cyclic and non-cyclic schedule. In the cyclic workforce schedule all employees have the same basic schedule but start with different offsets while in non-cyclic workforce schedule individual preferences of employees are taken into account. Due to the complexity of the problem with a lot of variables and constraints, a help of optimization software is necessary. Large number of various software packages is available to assist with rostering but they are usually targeted at a specific application area and cannot be easily transferred to another industry or company [2]. So it is always necessary to create a model for the given conditions. The scheduling or rostering process can be separated into 6 phases (sometimes not all of them are necessary) [2]:

- demand modeling – setting the number of staff necessary for each shift,
- days-off scheduling – description of the rest days for workers,
- shift scheduling - an assignment of the number of employees to each shift to meet demand,
- line of work construction – work schedule for each staff member,
- task assignment – if necessary, assignment of task to the shifts,
- staff assignment - assignment of individual staff to the lines of work.

The shift scheduling problems might be also divided on the basis of the days or hours – some are aimed at 5 workdays and 2 consecutive off-days, some at 5-7 days with limited workdays (airline crew, bus crew, call centers, restaurants staff etc.) and some at 7 workdays operating 24 hours a day like hospitals, manufacturing plants or retail stores [4]. Sometimes also the type of the work contract (full-time, part-time) is important to create a roster [7].

[1] University of Economics, Dpt. of Econometrics, W.Churchill Sq. 4, 130 67 Prague 3, kuncovam@vse.cz
[2] University of Economics, Dpt. of Econometrics, W.Churchill Sq. 4, 130 67 Prague 3, veronika.pickova@seznam.cz

In this paper we describe the construction of the system for the employees' non-cyclic shifts schedule on the basis of ILP solution with respect to all given real restrictions and requirements of employees and employer. The solution must be user-friendly for the managers, must speed up the shifts and rosters planning and must take into account various real-life situations. Firstly the model was created in MS Excel but because of the rising number of constraints, the help of LINGO software was necessary (for the model see [5]). Because of the possibilities and requirements of the company (one of the multiplex of the Cinema City Czech Republic) managers prefer the system to be prepared in MS Excel environment. They need it as a complex system where they can set the input requirements for one week and afterwards the weekly schedule is created via optimization model. That is why MS Excel macros were used to be able to operate it by managers without the knowledge of the model creation or the solver.

2 Data and methods

The company which the system was prepared for belongs to the net of the multiplex Cinema City in the Czech Republic. The system of the rostering for the shifts in the given company was not based on any software solution. All employees at first had to select shift and write it into a one week roster, than again all employees had to select their second shift in the given week etc. so as to cover all shifts. A manager was responsible for the correctly fulfilled form and for keeping the roster by employees. He/she must control if all shifts were covered in required capacity, whether no employee had two successive shifts in one day or that total number of working hours was not higher in one week than it was set in the employment contract. It was also necessary to change the roster in case of illness or absence of an employee. So the main task was first to create a model to ease the assignment of employees and second to prepare the assignment form and the optimization in MS Excel.

Data for the model were following: the roster concerns the position of barman where 10 employees might work. We call them person B_1, .., B_{10}. It is necessary to cover all shifts in one week – description of shifts with working hours is in Table 1.

day	Shift No.	Shift Name	Working hours	Number of hours	Min. n. of employees
Monday	1	MO1	11:00-15:00	4	1
	2	MO2	15:00-20:00	5	1
	3	MO3	20:00-01:00	5	2
Tuesday	4	TU1	11:00-15:00	4	1
	5	TU2	15:00-20:00	5	1
	6	TU3	20:00-01:00	5	2
Wednesday	7	WE1	11:00-15:00	4	1
	8	WE2	15:00-20:00	5	1
	9	WE3	20:00-01:00	5	2
Thursday	10	TH1	11:00-15:00	4	1
	11	TH2	15:00-20:00	5	1
	12	TH3	20:00-01:00	5	2
Friday	13	FR1	11:00-15:00	4	1
	14	FR2	15:00-20:00	5	1
	15	FR3	20:00-01:00	5	2
Saturday	16	SA1	9:00-15:00	6	1
	17	SA2	15:00-22:00	7	2
Sunday	18	SU1	9:00-15:00	6	1
	19	SU2	15:00-22:00	7	2

Table 1 Description of shifts in a week

Minimum number for employees for each shift is usually 1 but afterwards the managers defined that it is valid except of the last shift each day where 2 employees are necessary. Also each of the 10 employees must have at least 1 shift assigned (with the exception when the employee cannot be at work the given week). Each employee can only choose a day when he/she can/cannot be at work (they usually work for a part-time not for the full-time).

During weekend any of the employee cannot work in two consecutive days (so the combination Friday+Saturday or Saturday+Sunday is not allowed).

First of all it was necessary to decide how to transfer the information about the days when the employee can be at work into shifts and variables. We decided to create a table in MS Excel with yes/no possibility to choose for each employee and each day and afterwards to use IF function in MS Excel to change days into shifts and to create a table with input data (requirements of the employee, c_{ij} values). Also the cells for the managers' requirements should be defined in MS Excel. When we know we have 19 shifts in 7 days (Table 1) we are able to define which combination is not allowed (more than 1 shift in one day and more than one shift on weekend).

According to the previous information we define the mathematical model for selection of roster of all employees to shifts in one week. We suppose the set of employees $B = \{B_1, B_2, \dots, B_m\}, i = \{1, 2, \dots, m\}$, $m = 10$ and the set of shifts $S = \{S_1, S_2, \dots, S_n\}, j = \{1, 2, \dots, n\}$, $n = 19$. The aim is to assign the employees to the shifts so as to cover all shifts and with respect to other conditions given by managers of the company.

Mathematical model of the problem can be formulated as follows:

$$z = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} \rightarrow min \tag{1}$$

Subject to

$$x_{ij} \leq c_{ij} \quad i = 1, \dots, m, \qquad j = 1, \dots, n, \tag{2}$$

$$\sum_{j=1}^{n} x_{ij} \geq a_i, \quad i = 1, 2, \dots, m, \tag{3}$$

$$\sum_{i=1}^{m} x_{ij} = d_j, \quad j = 1, 2, \dots, n, \tag{4}$$

$$\sum_{j=k}^{k+2} x_{ij} \leq 1, \quad i = 1, 2, \dots, m, \quad k = 3p + 1, \qquad p = 0, 1, .., v - 1 \tag{5}$$

$$\sum_{j=k}^{k+1} x_{ij} \leq 1, \quad i = 1, 2, \dots, m, \quad k = n - (qn + 1), n - (qn + 3), \tag{6}$$

$$q = 0, 1, \dots, w - 1$$

$$\sum_{j=1}^{n} x_{ij} \geq r_j, \quad i = 1, 2, \dots, m, \tag{7}$$

$$x_{ij} \in \{0,1\}, \quad i = 1, \dots, m, \qquad j = 1, \dots, n \tag{8}$$

when

z	…	objective function minimizing number of employees on shifts (1),
x_{ij}	…	0-1 variable where:
		$x_{ij} = 1$ when employee *i* is assigned to shift *j*,
		$x_{ij} = 0$ otherwise,
c_{ij}	…	0-1 value where:
		$c_{ij} = 1$ when employee *i* indicated that could be on the shift *j*,
		$c_{ij} = 0$ otherwise,
a_i	…	minimal previously set number of assigned shifts for the *i*-th employee,
d_j	…	total number of employees necessary for the *j*-th shift,
m	…	number of employees (in our case $m = 10$),
n	…	number of shifts (in our case $n = 19$),
k	…	auxiliary index for shifts selection
p,q	…	auxiliary index for working/weekend day selection,
v	…	number of working days without weekend (in our case v=5),
w	…	number of weeks (in our case w=1)

r_j … min.required number of employees for the shift $_j$.

Descriptions of the constraints:

- an employee can be assigned on the shift only if he/she indicated in attendance form (Figure 1) that he/she can come to the given shift (2),
- an employee can be assigned to at least a_i shifts (3), where a_i is given by the manager,
- d_j employees are assigned to j-th shift (4), d_j requirements are set by the manager,
- an employee can be assigned only on max. one shift in a given day in a week (5) and max. one shift during Friday and Saturday and during weekend (6). Each day during a week has 3 shifts except of Saturday and Sunday where only 2 shifts are possible (so on Monday we have shifts 1,2 and 3, Tuesday shifts 4,5 and 6, etc.),
- at least r_j employees must be assigned to a shift j (7),
- employee i is/is not assigned to shift j (8).

3 Results and discussion

On the basis of the given data and requirements of the company managers firstly the attendance form had to be created for the employees (or managers) to indicate the days in a week the employee can be at work (Figure 1). The Excel sheet is interactive with yes and no possibilities for each employee and each day. The control cells were created to watch whether at least one employee selected "yes" in each day. If not, it is not possible to continue with the optimization and the table must be corrected. Afterwards a MS Excel macro was prepared to fill the input data table for 19 shifts with 0-1 values (1 when the employee can be on the shift, 0 when cannot be there). Then the optimization model for 4 employees only with respect to their requirements and with respect to 1 employee per shift was created using MS Excel solver. The conditions (5) and (6) were realized in MS Excel sheet by summarization of daily shifts for one employee but cannot be added as a constraints to the model as the limit 100 constraints was reached. In this sense the model creation in MS Excel is easy but the solver limits did not let us to create the whole model. One of the solutions of the test example is on Figure 2 (for this article the tables were translated into English but the add-in application for the company was prepared in Czech language only). We can see, that each shift is covered by 1 employee and also that each employee has at least one shift during the week.

Name	code	MO	TU	WE	TH	FR	SA	SU
	B1	yes	no	yes	no	no	no	yes
	B2	no	yes	no	yes	yes	yes	no
	B3	no	yes	yes	yes	yes	yes	no
	B4	yes	no	yes	no	yes	no	no
	B5	no	yes	yes	yes	no	no	no
	B6	no	yes	no	no	yes	yes	yes
	B7	yes	yes	yes	no	no	no	yes
	B8	yes	no	no	yes	yes	yes	no
	B9	no	no	no	no	no	yes	yes
	B10	no	yes	no	yes	no	no	no
		ok	ok	ok	ok	ok	ok	ok

CONTROL OK

Figure 1 Table for the employees (B1..B10) to indicate whether can ("yes") or cannot ("no") be the given day at work (MO=Monday, TU=Tuesday, WE=Wednesday, TH=Thursday, FR=Friday, SA=Saturday, SU=Sunday)

	MO 1	MO 2	MO 3	TU1	TU2	TU3	WE1	WE2	WE3	TH1	TH2	TH3	FR1	FR2	FR3	SA1	SA2	SU1	SU2
B1	0	0	1	0	0	1	0	0	1	0	1	0	0	0	1	0	0	0	1
B2	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0
B3	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
B4	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0
No of workers	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Required	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Length	4	5	5	4	5	5	4	5	5	4	5	5	4	5	5	6	7	6	7

Control	MO	TU	WE	TH	FR	SA	SU	FR+SA	SA+SU
B1	1	1	1	1	1	0	1	1	1
B2	1	1	1	1	1	0	0	1	0
B3	1	1	1	0	0	1	0	1	1
B4	0	0	0	1	1	1	1	2	2

Figure 2 Table with results (green cells)

The problem arise because of the 10 employees that are taken into account and the number of constraints given by the possibilities of workers to be at work on a given day. In this situation it would be necessary to separate the model into two parts (for example week days and weekend) to be able to solve it in MS Excel or to use a combination of MS Excel with outside solver (as it was necessary in our case for other conditions mentioned). We decided to use LINGO solver. As the managers after the first solution specified that it is usual to have at least 2 employees on the last shifts every day and as they prefer the MS Excel sheets for them to be able to operate with the model, the add-in application in MS Excel using LINGO solver was prepared.

We added the row with the number of required employees for each shift (called "required") and the column for minimum number of shifts for each employee in a week (called "min.shifts"). All the constraints (1), (2), .., (7) were rewritten to LINGO solver. One of the solutions is on Figure 3 (requirements are in last row and column in orange cells).

	MO 1	MO 2	MO 3	TU1	TU2	TU3	WE1	WE2	WE3	TH1	TH2	TH3	FR1	FR2	FR3	SA1	SA2	SU1	SU2	No of shifts	min. shifts
B1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1
B2	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	3	1
B3	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	3	1
B4	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	3	1
B5	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	3	1
B6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	2	1
B7	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	3	1
B8	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	3	1
B9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2	1
B10	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	2	1
No of workers	1	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1	2	1	2		
Required	1	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1	2	1	2		

Figure 3 Table with results (green cells) with previously set parameters (orange cells)

The last requirement of managers was to add working hours to be able to set maximum working hours for each employee at the given week. The model in MS Excel was enlarged by the columns with the assigned working hours (called "hours" for each employee and "length" for each shift) and with maximal possible working hours (called "max.hours"). In MS Excel the change of the model was not complicated – only the function SUMPRODUCT was added to calculate the number of working hours for each employee. The constraints describing, that these values must be lower than the maximums set, were added into LINGO solver. The solution when the maximum working hours for each employee were set to 17 hours is on Figure 4. We also see number of shifts for each employee during the given week.

	MO 1	MO 2	MO 3	TU1	TU2	TU3	WE1	WE2	WE3	TH1	TH2	TH3	FR1	FR2	FR3	SA1	SA2	SU1	SU2	No of shifts	min. shifts	hours	max. hours
B1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	3	1	17	17
B2	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	3	1	17	17
B3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	3	1	17	17
B4	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	3	1	13	17
B5	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	2	1	8	17
B6	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	3	1	16	17
B7	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	1	15	17
B8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	3	1	16	17
B9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	7	17
B10	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	1	9	17
No of workers	1	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1	2	1	2				
Required	1	1	2	1	1	2	1	1	2	1	1	2	1	1	2	1	2	1	2				
Length	4	5	5	4	5	5	4	5	5	4	5	5	4	5	5	6	7	6	7				

Figure 4 Table with results (white and green cells) with previously set parameters (orange cells) including working hours

The model helps the managers to quickly create a roster for the selected week when the information about the employees' availability is known. The solution is easy to obtain and the inputs and results are easy to understand. The only problem is that this size of the model exceeds the MS Excel possibilities as it has usually max. 200 variables and 100 conditions while in the given case we have 190 variables and between 290-319 constraints. There are other possibilities how to change the model: look for the alternative solution (additional constraint for one of the non-zero variable to be zero), exclude the cooperation of two selected employees, concentrate selected shifts into neighboring days, change the model into goal programming with given goals, simulate the process with different employees quality etc. – some of them were mentioned and solved in [5].

4 Conclusions

Shift scheduling and rostering belongs to the problems that can be solved via optimization models. The purpose of this study was to create an application in MS Excel for the non-cyclic shift scheduling problem for one week that is easy to use for the managers. The main aim was the minimization the number of staff required. Employees should be assigned to the shifts based on the days they have selected as their working days in the given week. Also the minimal number of workers per each shift and the maximal working hours weekly per each worker must be kept. The application has a control functions signalizing that some conditions are not fulfilled (a day is not covered by any employee, maximum working hours are not set, minimum requirements for the shift is not set). The optimization model for 10 employees and 19 shifts with fixed inputs is solved by MS Excel with LINGO solver combination as the number of constraints exceeds the MS Excel settings. The combination of MS Excel and LINGO provides the results quickly, using macros in MS Excel simplifies the visualization and insertion of the inputs and also the visualization and understanding of the results. On the other hand it is not possible to use MS Excel only (as it was required by the managers) as the model should be separated into two or three models to decrease the number of constraints. Future development of the model can take into account other requirements of employees or managers. Although the whole model is not complicated (from the operations research point of view), the managers were surprised by the quick response of the software and by the quality of the results as it reduced their time with the shift scheduling (it usually takes several hours from the employees' filling form of the possible working days till the final schedule created by manager).

Acknowledgements

This work was supported by grant IGA 57/2017 of the Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Azaiez, M.N., Al Sharif, S.S.: A 0-1 goal programming model for nurse scheduling. *Computers & Operations Research* **32** (2005), 491–507.

[2] Dowling, D., Krishnamoorthy, M., Mackenzie, H., Sier, D.: Staff rostering at a large international airport. *Annals of Operations Research* **72**(1997)125–147.

[3] Ernst, A.T., Jiang, H., Krishnamoorthy, M., Sier, D.: Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research* **153** (2004) 3–27.

[4] Musliu, N., Gartner, J., Slany, W.: Efficient generation of rotating workforce schedules. *Discrete Applied Mathematics* 118/1-2 (2002), 85-98.

[5] Picková, V.: Přiřazování pracovníků na směnu: aplikace vybraných metod operačního výzkumu. Master theses, University of Economics, Prague, 2016.

[6] Skočdopolová, V.: Optimisation of University Timetable via Complex Goal Programming Model. In: *Proceedings of the 33rd Mathematical Methods in Economics Conference*, Cheb, Czech Republic, 2015, 725-730.

[7] Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E.: Personnel scheduling: A literature review. *European Journal of Operational Research* **226** (2013) 367–385.

Struggle with Curse of Dimensionality in Robust Emergency System Design

Marek Kvet [1], Jaroslav Janáček[2]

Abstract. Emergency service system design is determined by deployment of limited number of service centers at positions from a given set of possible locations. The objective is to minimize the sum of weighted distances from particular system users to the nearest located service center. The robust service system is designed to comply with specified scenarios by minimizing the maximal value of the objective functions corresponding to the particular scenarios, which represent consequences of random failures in the road network. The events may be due to congestion, disruptions or blockages. The min-max link-up constraints and the cardinality of the scenario set represent an undesirable burden in any solving process used for design solution. Within this paper, we present several attempts to mastering the above drawback and provide the reader with a comparison of suggested approximate methods for the robust emergency system design.

Keywords: Emergency system design, robustness, approximate solving methods

JEL Classification: C61
AMS Classification: 90C06, 90C10, 90C27

1 Introduction

Optimization and methods of operational research play a very important role in many areas of human life. This paper is focused on special class of discrete network location problems, which are solved under uncertainty following from randomly occurring failures in the transportation network, through which the associated service is provided [2], [9], [10]. Studied approaches are applied on emergency service system design problems.

The emergency system design problem is a challenging task for both system designer and operational researcher. As the first one searches for a tool, which enables to obtain service center deployment satisfying future demands of the system users in case of emergency, the second one faces the necessity of completing the solving tool. Emergency service system efficiency is considerably influenced by deployment of the service centers, which send emergency vehicles to satisfy demands on service at system users' locations. The number of service providing centers must be limited due to economic and technological restrictions. If the quality characteristic of the design corresponds to service accessibility of an average user, then the emergency service system design can be tackled as the weighted *p*-median problem, which was studied in [1], [3], [4] and many others from the points of effective exact and approximate solving techniques. We also assume that each service center has enough capacity to serve all assigned users. Thus, each system user is served only from the nearest located service center.

When the emergency service system is designed, the designer must take into account that the traversing time between a service center and the affected user might be impacted by various random events caused by weather or traffic. In other words, the system resistance to such critical events is demanded. Most of available approaches to increasing the system resistance [5], [6], [7], [8] are based on making the design resistant to possible failure scenarios, which can appear in the associated transportation network as a consequence of random failures due to congestion, disruptions or blockages. Then, an individual scenario is characterized by particular time distances between the users' locations and possible service center locations.

Let us focus now on the objective function value. The most commonly used objective function in the weighted *p*-median problem consists in minimizing the time service accessibility for an average user [1], [4]. If we want to make the system resistant to randomly occurring detrimental events, then the objective function requires a different approach. A robust service system design has to comply with all specified scenarios. The usual way of taking

[1] Ing. Marek Kvet, PhD., University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovak Republic, e-mail: marek.kvet@fri.uniza.sk
[2] Prof. RNDr. Jaroslav Janáček, CSc., University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovak Republic, e-mail: jaroslav.janacek@fri.uniza.sk

into account all scenarios is based on minimizing the maximal objective function of the individual instances corresponding with particular scenarios [9], [10]. It means that the min-sum objective function used in the standard weighted *p*-median problem is replaced here by the min-max quality criterion. The min-max link-up constraints and the cardinality of the scenario set represent an undesirable burden in any integer programming problem due to bad convergence of the branch-and-bound method inside most available IP-solvers. Thus, complementary approximate approaches to the robustness constitute a big challenge to family of operational researchers and professionals in applied informatics. In this paper, we present several attempts to mastering the above drawback and provide the reader with a comparison of suggested approximate methods for the robust emergency system design. Presented approaches are based on Lagrangean relaxation [5], fuzzy set theory [7], [11], special supplementary scenarios [6] or other advanced iterative algorithms [8].

The remainder of this paper is organized as follows: Section 2 is devoted to the description of original min-max robust design of emergency system, in which all scenarios are taken into account. Suggested approximate approaches based on several modelling techniques are explained in Section 3. The fourth Section contains the overview of performed numerical experiments and yields brief comparative analysis of designed service center deployments. The results and findings are summarized in Section 5.

2 Standard approach to the robust emergency system design problem

The robust emergency system design problem can be described using the following denotations. Let symbols J and I denote the set of users' locations and the set of possible service center locations respectively. Symbol b_j denotes the number of users sharing the location j and p denotes the maximal number of service centers to be chosen from the set I. Let symbol U denote the set of possible failure scenarios. Particular scenario may correspond with some situation in the underlying transportation network, e.g. weather, congestions and other critical events influencing travelling times.

The usual objective of the emergency system design problem is to minimize system disutility of an average user. The value of the user's disutility is determined by mutual positions of the user location and the location of the center providing the service. Disutility following from the distance between locations i and j under a specific scenario $u \in U$ is denoted here as d_{iju}. We assume that the user's disutility grows with increasing distance between the user and the service center. Especially, the disutility perceived by a user of an emergency system may correspond to estimated travelling time in minutes. In this paper, we consider that each value of d_{iju} is integer and less than or equal to the maximal value D_{max}. If we want to make the system resistant to randomly occurring detrimental events, then the objective function of the system design turns into minimizing the maximal objective function of the individual instances corresponding with particular scenarios.

Complexity of location problems with limited number of facilities to be deployed and the necessity to solve large instances of the problem led to searching for a suitable algorithm. It was found that the radial formulation of the problem could considerably accelerate the associated solving process [3], [4]. As this concept proved to be a suitable tool, we decided to apply the radial formulation also on the robust emergency system design.

To complete the associated mathematical model of the robust emergency system design problem, several decision variables need to be introduced. The variable $y_i \in \{0,1\}$ models the decision on service center location at the location $i \in I$. The variable takes the value of 1 if a service center is located at i and it takes the value of 0 otherwise. In the robust problem formulation, the variable h denotes the upper bound of the objective function issues over the set U of scenarios. Let us define $v = D_{max} - 1$. Further, auxiliary zero-one variables x_{jsu} for $s = 0 \ldots v$ and $u \in U$ are introduced to complete the radial model. The variable x_{jsu} takes the value of 1, if the disutility of the user at $j \in J$ from the nearest located center under the scenario $u \in U$ is greater than s and it takes the value of 0 otherwise. Then the expression $x_{j0u} + x_{j1u} + \ldots + x_{jvu}$ constitutes the value of disutility d_{ju*} from the user location j to the nearest located service center under the scenario $u \in U$. Let us introduce a zero-one constant a_{iju}^s under the scenario $u \in U$ for each triple $[i, j, s]$, where $i \in I$, $j \in J$, $s \in [0..v]$. The constant a_{iju}^s is equal to 1, if the disutility d_{iju} between the user location j and the possible center location i is less than or equal to s, otherwise a_{iju}^s is equal to 0. Then the model, in which the maximum of the objective function values over the set U is minimized, follows.

$$\textit{Minimize} \qquad h \tag{1}$$

$$\textit{Subject to}: \quad x_{jsu} + \sum_{i \in I} a_{iju}^s y_i \geq 1 \quad \textit{for } j \in J, s = 0, 1, \ldots, v, \; u \in U \tag{2}$$

$$\sum_{i \in I} y_i \leq p \tag{3}$$

$$\sum_{j \in J} b_j \sum_{s=0}^{v} x_{jsu} \leq h \quad \text{for } u \in U \tag{4}$$

$$y_i \in \{0, 1\} \quad \text{for } i \in I \tag{5}$$

$$x_{jsu} \geq 0 \quad \text{for } j \in J, s = 0, 1, \ldots, v, \; u \in U \tag{6}$$

$$h \geq 0 \tag{7}$$

In this model, the objective function (1) represented by single variable h gives the upper bound of all objective function values over the individual scenarios. The constraints (2) ensure that the variables x_{jsu} are allowed to take the value of 0, if at least one center is located in radius s from the user location j and constraint (3) limits the number of located service centers by p. The link-up constraints (4) ensure that each perceived disutility is less than or equal to the upper bound h. As the obligatory constraints (6) are concerned, only values zero and one are expected in any feasible solution, but it can be seen that the model has integrality property restricted to the variables x_{jsu}. It follows that the relevant values of x_{jsu} in the optimal solution will be equal to one or zero without imposing binary constraints upon these variables.

Despite the fact that the radial formulation may accelerate the associated solving process [3], [4], the min-max link-up constraints (4) represent an undesirable burden in any integer programming problem due to bad convergence of the branch-and-bound method and make the robust emergency system design problem hard to solve. Therefore, many approximate approaches have been developed to avoid mentioned obstacle.

3 Approximate modelling methods

The main goal of this chapter is to provide the readers with a brief overview of possible ways to comply with bad convergence of the branch-and-bound method applied on the model (1) – (7) with the link-up constraints (4) linking the individual scenario objective functions up to their common upper bound. We present here several approaches based on the radial formulation, which have been recently developed.

The first suggested approach follows the Lagrangean relaxation of the link-up constraints (4). Each of these constraints is associated with a non-negative Lagrangean multiplier λ_u and the sub-gradient method is used to set up suitable values of the multipliers. The original problem (1) – (7) turns into minimization of (8) subject to (2), (3), (5), (6) and (7).

$$\textit{Minimize} \qquad h + \sum_{u \in U} \lambda_u \left(\sum_{j \in J} \sum_{s=0}^{v} b_j x_{jsu} - h \right) = h \left(1 - \sum_{u \in U} \lambda_u \right) + \sum_{u \in U} \lambda_u \sum_{j \in J} \sum_{s=0}^{v} b_j x_{jsu} \tag{8}$$

Obviously, the problem (8), (2), (3), (5), (6) and (7) has solution only for such setting of Lagrangean multipliers, where their sum is less than or equal to one. Having the optimal solution of the relaxed problem with arbitrary multipliers meeting the stated rule, the value of the optimal solution yields a lower bound of the optimal solution of the original problem (1) - (7). If the sum of the Lagrangean multipliers is less than one, then the optimal value of h would equal to zero and it does not represents the upper bound at all. That is why we restrict ourselves on such setting of multipliers, where the sum equals to one exactly. We try to reach or at least to approximate the optimal solution of the original problem (1) - (7) by an iterative algorithm, where the Lagrangean multipliers λ_u are recomputed in each iteration. This algorithm was introduced in [5], where also other details are discussed.

The rest of presented approaches are based on the idea of replacing the set U of individual scenarios by one supplementary scenario. After this assumption, the original model (1) – (7) can be simplified to the form (9) – (13).

$$\textit{Minimize} \qquad \sum_{j \in J} b_j \sum_{s=0}^{v} x_{js} \tag{9}$$

$$\textit{Subject to}: \quad x_{js} + \sum_{i \in I} a_{ij}^{s} y_i \geq 1 \quad \text{for } j \in J, s = 0, 1, \ldots, v \tag{10}$$

$$\sum_{i \in I} y_i \leq p \tag{11}$$

$$y_i \in \{0, 1\} \quad for\ i \in I \tag{12}$$

$$x_{js} \geq 0 \quad for\ j \in J, s = 0, 1, \ldots, v \tag{13}$$

All decision variables and structural constraints take the same meaning as before. The objective function (9) minimizes the disutility perceived by an average user, what can be achieved by minimizing the sum of disutility values perceived by all system users. The only difference consists in the fact, that the indexes u over the set U of scenarios used in the variables and constraints of the previous model (1) – (7) are not necessary.

Generally, the constant a_{ij}^s for $i \in I$, $j \in J$, $s \in [0..v]$ is equal to 1, if the disutility $\underline{d}_{ij}$ perceived by any user located at j from the possible center location i is less than or equal to s, otherwise $a_{ij}^s = 0$. The disutility $\underline{d}_{ij}$ may be defined in many different ways.

The fuzzy approach [7] does not deal with the set of given crisp scenarios as proposed in the model (1) – (7), but it describes the individual uncertain values, e.g. the value of perceived disutility, by a range of possible values together with a measure of relevance of the individual values from the range. The measure is called membership function and the function is defined on the whole set of real numbers and maps this definition range on the interval [0, 1]. The value of one is assigned to the real values which belong to the fuzzy value at the highest level and the value of zero is intended for real values outside the range of the fuzzy value. A particular fuzzy set defined on real numbers membership function of which satisfies given conditions is called fuzzy number. We restrict ourselves to a special type of the fuzzy sets intended for expressing that a processed uncertain disutility value is greater than or equal to a given real value $d_{ij\text{-}min}$ with tolerance $d_{ij\text{-}Max} - d_{ij\text{-}min}$. Here, $d_{ij\text{-}min} = min\{d_{iju}: u \in U\}$. Similarly, $d_{ij\text{-}Max} = max\{d_{iju}: u \in U\}$.

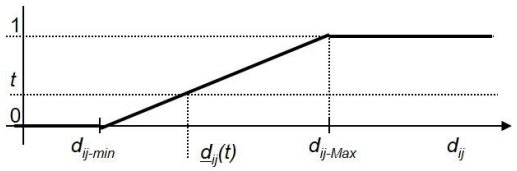


Figure 1 The membership function of a fuzzy set $\underline{d}_{ij}$

Then, the constant a_{ij}^s must be defined for each pair $[i, j]$ and given level of satisfaction $t \in [0, 1]$. The constant $a_{ij}^s(t)$ is equal to 1, if the disutility $\underline{d}_{ij}(t)$ between the user location j and the possible center location i is less than or equal to s, otherwise $a_{ij}^s(t) = 0$. After these preliminaries, we can solve the model (9) – (13) with fuzzy coefficients by the iterative Tanako-Asai's approach, in which we use the bisection method to search for the highest level of satisfaction t^*, for which the problem has a feasible solution. The achieved results are reported in [7].

The disadvantage of previously introduced approach consists in the fact that it does not process the complete set U in a sufficient way. The minimum and maximum of each disutility over the set U does not bring any information about other scenarios, which should be also considered. Therefore, we have suggested another approach.

This new approach is based on a convex combination of individual scenarios. Similarly to the Lagranean relaxation described above, the set of nonnegative multipliers $\lambda_u \geq 0$ for $u \in U$ is used to define the resulting value of each disutility $\underline{d}_{ij}$ perceived by any user located at j from the possible center location i. It must be noted that the coefficients λ_u have to satisfy the condition that their sum equals to one. These multipliers can also represent the probability distribution of the individual scenarios. Then, the disutility values $\underline{d}_{ij}$ can be expressed as follows.

$$\underline{d}_{ij}(\lambda) = \sum_{u \in U} \lambda_u d_{iju} \quad for\ i \in I,\ j \in J \tag{14}$$

The values of $\underline{d}_{ij}$ depend on the settings of multipliers λ_u. After these preliminaries, the covering constants a_{ij}^s are computed and the radial model (9) – (13) is iteratively solved. In each iteration, the multipliers λ_u are redefined based on the obtained solution according to certain rules [8]. The radial model is being solved, while better solution of the robust emergency system design keeps being obtained or until other stopping criteria are met.

The last approach [6] was developed as an alternative to avoid iterative solving of the radial model (9) - (13) and to overcome the necessity of adjusting the model parameters. The values of $\underline{d}_{ij}$ represent here the worst situation, which may occur in the associated transportation network. In this method, $\underline{d}_{ij} = max\{d_{iju}: u \in U\}$.

The main goal of this paper is to experimentally compare all suggested approaches and to decide, which of them represents the most suitable tool for replacing the exact method with the min-max link-up constraints.

4 Computational study

To compare presented approaches to robust emergency system design from the points of computational time demands and the solution accuracy, we performed the series of numerical experiments. Here, the optimization software FICO Xpress 8.0 (64-bit, release 2016) was used and the experiments were run on a PC equipped with the Intel® Core™ i7 5500U processor with the parameters: 2.4 GHz and 16 GB RAM.

The used benchmarks were derived from the real emergency health care system, which was originally implemented in seven regions of Slovak Republic. For each self-governing region, i.e. Bratislava (BA), Banská Bystrica (BB), Košice (KE), Nitra (NR), Trenčín (TN), Trnava (TT) and Žilina (ZA), all cities and villages with corresponding number of inhabitants b_j were taken into accountn. The coefficients b_j were rounded to hundreds. In the benchmarks, the set of communities represents both the set J of users' locations and the set I of possible center locations as well. The cardinalities of these sets vary from 87 to 664 according to the considered region. The number p of located centers varies from 9 to 67. The network distance from a user to the nearest located center was taken as the user´s disutility.

An individual experiment was organized so that the exact model (1) - (7) was used to obtain the optimal solution of the robust emergency system design problem and then, the approximate methods were applied to study their characteristics. Due to the lack of common benchmarks for study of robustness, the scenarios used in our computational study were created in the following way. We selected 25 percent of matrix rows so that these rows correspond to the biggest cities concerning the number of system users. Then we chose randomly from 5 to 15 rows from the selected ones and the associated disutility values in the chosen rows were multiplied by the randomly chosen constant from the range 2, 3 and 4. The rows, which were not chosen by this random process, stay unchanged. This way, 10 different scenarios were generated for each self-governing region. The scenarios represent the consequence of fatal detrimental events, when some time-distances are several times elongated.

The achieved results of our computational study are reported in Table 1, in which all presented approaches to the robust emergency system design are compared. The computational times in seconds are given in the columns denoted by CT. Since the approximate methods use different objective functions than used in the model (1) – (7), the resulting system designs were compared from the point of robust min-max objective function value h, which was evaluated for each design according to the expression (15).

$$h = \max\left\{\sum_{j\in J} b_j \min\left\{d_{iju} : i \in I,\ y_i = 1\right\} : u \in U\right\} \tag{15}$$

Finally, the value of *Dif* is reported for each approximate method. This value expresses the difference of the resulting objective function value from the optimal one in percentage, where the objective function value of the model (1) – (7) was taken as the base.

	$\|I\|$	p	**Standard (1)-(7)**		**Lagrange relaxation**			**Fuzzy approach**			**Convex combination**			**Maximal scenario**		
			CT	***h***	**CT**	***h***	***Dif***	**CT**	***h***	***Dif***	**CT**	***h***	***Dif***	**CT**	***h***	***Dif***
BA	87	9	25.4	25417	39.3	26197	3.1	2.9	27014	6.3	13.1	25417	0.0	0.1	27183	6.9
BB	515	52	953.7	18549	166.0	18859	1.7	44.2	19458	4.9	53.6	19056	2.7	3.7	19276	3.9
KE	460	46	1150.5	21286	211.4	21522	1.1	35.2	22872	7.5	36.5	21717	2.0	0.5	22790	7.1
NR	350	35	1862.1	24193	112.2	24440	1.0	25.8	24849	2.7	53.8	24372	0.7	0.8	25137	3.9
TN	276	28	267.0	17524	41.7	17914	2.2	16.0	17883	2.0	28.1	17605	0.5	3.8	18222	4.0
TT	249	25	394.2	20558	150.8	20560	0.0	10.2	21639	5.3	29.2	20859	1.5	1.2	22160	7.8
ZA	315	32	1132.3	23004	35.8	23183	0.8	13.9	23655	2.8	39.5	23156	0.7	0.3	23808	3.5

Table 1 Results of numerical experiments for the self-governing regions of Slovakia.

The obtained results have proved our expectations. The approximate methods perform in order faster than the original standard exact approach, which is formulated by the model (1) – (7). The difference is caused by the fact, that the approximate approaches do not process all scenarios and avoid introducing the min-max link-up constraints, which cause bad convergence of the branch-and-bound method.

As concerns the solution accuracy, the results obtained in this computational study indicate that the most suitable solving tool is the convex approach, because it takes into account the complete set U of scenarios and does not reflect only to some selected values of disutility. Therefore, we will continue in this algorithm, we will try to adjust its parameters to make the algorithm more efficient and accurate.

5 Conclusions

This paper was focused on mastering large instances of the robust emergency system design problem using commercial IP-solver. The robustness follows the idea of making the system resistant to various randomly occurring catastrophic events. The original approach with the min-max objective function value proved to be extremely time consuming due to the fact, that the min-max link-up constraints cause bad convergence of the branch-and-bound method inside most IP-solvers. This situation occurs also in cases, in which the radial formulation of the problem is used. This obstacle can be overcome by approximate solving methods, which are based on several different approaches. In this paper, we have performed a computational study to compare Lagrangean relaxation of the link-up constraints to fuzzy approach and to other modelling techniques, which create a convex combination of individual scenarios or replace the set of scenarios by one subsidiary scenario representing the worst situation that may occur in the associated transportation network. As concerns the accuracy of the resulting solution, it can be observed that the approximate methods are very satisfactory. Thus, we can conclude that we have presented many useful tools for the robust emergency system design problem, which can be easily implemented using common commercial optimization software.

The future research in this field could be aimed at finding relevant scenarios, which can significantly affect the performance of emergency service system. Since the robust design can significantly differ from the basic one, it would be useful to find such method, which allows changing only limited number of service center locations in comparison with the standard solution.

Acknowledgement

This work was supported by the research grants VEGA 1/0518/15 “Resilient rescue systems with uncertain accessibility of service”, VEGA 1/0463/16 “Economically efficient charging infrastructure deployment for electric vehicles in smart cities and communities”, APVV-15-0179 “Reliability of emergency systems on infrastructure with uncertain functionality of critical elements”.

References

[1] Avella, P., Sassano, A., Vasil'ev, I..: *Computational study of large scale p-median problems*. In: Mathematical Programming 109, pp. 89-114, 2007.

[2] Correia, I., Saldanha da Gama, F.: Facility locations under uncertainty. *Location Science*, eds. Laporte, Nikel, Saldanha da Gama, pp. 177-203, 2015.

[3] García, S., Labbé, M., Marín, A.: *Solving large p-median problems with a radius formulation*. In: INFORMS Journal on Computing 23 (4), pp. 546-556, 2011.

[4] Janáček, J.: *Approximate Covering Models of Location Problems*. In: Lecture Notes in Management Science: Proceedings of the 1st International Conference ICAOR ´08, Vol. 1, Sept. 2008, Yerevan, Armenia, pp. 53-61, ISSN 2008-0050, 2008.

[5] Janáček, J., Kvet, M.: Designing a Robust Emergency Service System by Lagrangean Relaxation. In Mathematical Methods in Economics 2016, Liberec, 6.-9.9. 2016, pp. 349-353, 2016.

[6] Janáček, J., Kvet, M.: Návrh robustního záchranného systému pomocí maximálního scénáře. In: Nové trendy v ekonometrii a operačním výzkumu, Bratislava: Ekonóm, pp. 55-59, 2016.

[7] Janáček, J., Kvet, M.: An Approach to Uncertainty via Scenarios and Fuzzy Values. In: *Croatian Operational Research Review* 8 (1), pp. 237–248, 2017.

[8] Kvet, M., Janáček, J.: Hill-Climbing Algorithm for Robust Emergency System Design with Return Preventing Constraints. In: 9th International Conference on Applied Economics: Contemporary Issues in Economy, 2017, Toruń, Poland, in print, 2017.

[9] Pan, Y., Du, Y., Wei, Z.: Reliable Facility System Design Subject to Edge Failures. *American Journal of Operations Research*, 4, pp. 164-172, 2014.

[10] Scaparra, M.P., Church, R.L.: Location Problems Under Disaster Events. In: Location Science, eds. Laporte, Nikel,Saldanha da Gama, pp. 623-642, 2015.

[11] Teodorovič, D., Vukadinovič, K.: Traffic control and transport planning: A fuzzy sets and neural networks approach. Boston: Kluwer Academic Publishers, 387 p., 1998.

Weak Orders for Intersecting Lorenz Curves: Analysis of income distribution in European countries

Tommaso Lando[1], Lucio Bertoli-Barsotti[2], Michaela Staníčková[3]

Abstract. The low growth of the EU performance has increased concerns regarding an increasing wage dispersion and income inequality, which hampers sustainable and inclusive economic development. In this context, the Lorenz curve can be used to define different ranking criteria for comparison of income distributions in terms of inequality, such as the Lorenz dominance and some weaker dominance relations that have been introduced more recently in the literature. In this paper, we review such methods and propose an empirical study with which we analyse and compare the income distributions of the 28 European Union member states (EU-28) plus other five European countries, on the basis of the dominance criteria discussed. The database for empirical analysis has been downloaded from the Eurostat's website – Income and living conditions statistics.

Keywords: Lorenz curve, income inequality, disparity, stochastic dominance, EU-28.

JEL Classification: C44, E24, O52.
AMS Classification: 62C, 91B14.

1 Introduction

In this study, the term "income" is referred to as household disposable income in a particular year, which consists of earnings, self-employment, capital income and public cash transfers; income taxes and social security contributions paid by households are deducted. The income of the household is attributed to each of its members, with an adjustment to reflect differences in needs for households of different sizes.

The Lorenz curve (LC) is a fundamental tool for comparison of income distributions in terms of inequality. The LC gives rise to a preorder, that is, the Lorenz dominance (LD), which is generally used to rank distributions based on their degree of inequality. In an economic framework, the LD is coherent with the Pigou-Dalton condition, i.e. the so called "principle of transfers". According to this principle, the higher of two non-intersecting Lorenz curves can be obtained from the lower one by an iteration of income transfers from "richer" to "poorer" individuals (the so called *elementary transfers* or *T-transforms*, see [8], p. 32), also called *progressive transfers*, [10]). For this reason, the "coherence" with the LD represents a fundamental property for all inequality (or concentration) measures.

However, many empirical studies revealed that Lorenz curves often intersect in the practice. It is not rare to find couples of distributions that cannot be ranked by the LD. In such cases, we can compare the intersecting distributions by relying on weaker orders of inequality, such as the *second-degree Lorenz dominance, 2-LD* [1]. The basic idea of the 2-LD is to move from the (first-degree) LD to a dominance relation of higher degree by cumulating LCs i) from the bottom or, ii), from the top, which gives rise to a Lorenz dominance of second degree. In the first case (i), this idea has been analysed in several works, related to the concept of third-degree inverse stochastic dominance [1,9]. These orderings emphasize the left tail of the distribution. Indeed, many authors (see e.g. [2,10]) agree that an elementary transfer should be more equalizing the "lower" it occurs in the distribution, that is, the principle of *aversion to downside inequality* [4]. On the other hand, in an economic context, a lot of attention is recently given to those variations occurring at the top of the income distribution [7]. Thus, the second idea (ii), introduced by Muliere and Scarsini [9] and more recently developed by Aaberge [1], emphasizes the right tail of the distribution in spite of the left one. More recently, Lando and Bertoli-Barsotti [6] introduced a new weak order, namely *disparity dominance*, which is based on the cumulated difference between the upper and the lower parts of the LC and it is aimed at emphasizing inequality in both tails of the distribution. Indeed, this approach attempts to combine the main features of approaches i) and ii) into a single preorder.

[1] University of Bergamo, via dei Caniana 2, Bergamo, Italy and VŠB-TU Ostrava, Faculty of Economics, Department of Finance, Sokolská třída 33, 702 00 Ostrava 1, Czech Republic e-mail: tommaso.lando@unibg.it
[2] University of Bergamo, via dei Caniana 2, Bergamo, Italy e-mail: lucio.bertoli-barsotti@unibg.it
[3] VŠB-TU Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 702 00 Ostrava 1, Czech Republic e-mail: michaela.stanickova@vsb.cz

In this paper, we analyse the income distributions of the EU-28 countries plus other five non-WU countries in the year 2015 with a twofold objective: i) compare the effectiveness of the weak preorders studied in terms of discriminating power and; ii) characterize mathematically the differences between the distributions of income of the EU-28 countries. In section 2, we introduce the notation and present the dominance relations used in our empirical analysis. In section 3 we present the data and describe the empirical results. In section 4 we summarize the results.

2 Methods

We recall that a preorder is a binary relation $\leq$ over a set S that is reflexive and transitive. In particular, observe that a preorder $\leq$ does not generally satisfy the antisymmetric property (that is, $a \leq b$ and $b \leq a$ does not necessarily imply $a = b$) and it is generally not total (that is, each pair a, b in S is not necessarily related by $\leq$).

In this paper, we relax the common assumption of non-negativity of the distribution, because our data in section 3 present also negative incomes. Although this feature does not hamper the methods discussed below, in what follows we shall discuss the effect of negative incomes on the fundamental tools for the measurement of inequality, such as the LC and the Gini index.

Let F be a distribution with positive and finite expectation μ_F (clearly, the mean cannot be zero or less than zero, otherwise the LC cannot be defined). The (generalized) inverse or quantile function of F is given by:

$$F^{-1}(p) = \inf\{z: F(z) \geq p\}, p \in (0,1). \tag{1}$$

The Lorenz curve L_F is defined as follows:

$$L_F(p) = \frac{1}{\mu_F}\int_0^p F^{-1}(t)dt\,, p \in (0,1). \tag{2}$$

The LC is a convex function. However, note that, differently by the usual definition of [5], the LC may be decreasing and negative because of negative incomes.
The Gini index is given by twice the area between the Lorenz curve and the 45° line:

$$\Gamma(F) = 1 - 2\int_0^1 L_F(t)dt. \tag{3}$$

Under the common assumption of non-negative incomes, the Gini index is typically normalized between 0 and 1. However, in presence of negative incomes, the LC can be negative, so that Γ can exceed 1. For this reason, Chen, Tien-Wang and Tong-Shieng [3] propose re-normalizing Γ by dividing it by $\Gamma_{max} = 1 - 2\int_0^h L_F(t)dt$, where $h = \sup\{p: L_F(p) \leq 0\}$. The resulting index is denoted here by $\Gamma^* = \Gamma/\Gamma_{max}$.
Let us also define the complementary curve $\bar{L}_F$, given by:

$$\bar{L}_F(p) = \frac{1}{\mu_F}\int_0^p F^{-1}(1-t)dt = 1 - L_F(1-p), p \in (0,1) \tag{4}$$

Actually, for a given percentage p, $L_F(p)$ represents the percentage of "total" possessed by the low $100p\%$ part of the distribution, while $\bar{L}_F(p)$ represents the percentage of "total" corresponding to the top $100p\%$ part of the distribution. From a geometrical point of view, $\bar{L}_F(p)$, increasing and concave, is the 180° rotation of $L_F(p)$, increasing and convex, with respect to the point $(0.5,0.5)$. Now we can introduce:

$$\Delta_F(t) = \bar{L}_F(t) - L_F(t), t \in [0,1/2]. \tag{5}$$

The interpretation of Δ_F is quite simple. As $\bar{L}_F(t) \geq L_F(t)\ \forall t \in [0,1/2]$, the difference between the Lorenz curves expresses the disparity between the "higher" and the "lower" parts of the distribution. In terms of income distributions, Δ_F equals the difference between the proportion of the society's overall wealth that is held by the society's top (rich) $100t\%$, and the proportion of the society's overall wealth that is held by the society's low (poor) $100t\%$.

It should be stressed, that the definitions below have been originally introduced only for the case of non-decreasing and non-negative LCs. However, there is no loss of generality if we apply them to our definition of the LC, which also account for the possibility of negative incomes.

Definition 1. We write $F \leq_L G$ if and only if $L_F(p) \geq L_G(p), \forall\, p \in (0,1)$.

When the LD is not fulfilled, i.e. when LCs intersect, we need to introduce some weaker criteria in order to obtain unambiguous rankings. Muliere and Scarsini [9] and Aaberge [1] suggest cumulating LCs from the left or from right: that is, attaching more weighting to low or top incomes. Such two parallel approaches give rise to the *upward* and *downward* 2-LD [1], defined as follows.

Definition 2. We say that G second-degree upward Lorenz dominates F, and write $F \leq_L^2 G$ if

$\int_0^t L_F(p)dp \geq \int_0^t L_G(p)dp$, $\forall t \in [0,1]$.

We say that G second-degree downward Lorenz dominates F, and write $F \leq_L^2 G$ iff any of the following equivalent conditions is true:

$\int_0^t \bar{L}_F(p)dp \leq \int_0^t \bar{L}_G(p)dp$, $\forall t \in [0,1]$

$\int_t^1 1 - L_F(p)dp \leq \int_t^1 1 - L_G(p)dp$, $\forall t \in [0,1]$.

Observe that $F \leq_L G$ implies $F \leq_L^2 G$ and $F \leq_{\bar{L}}^2 G$, but the converse is not necessarily true, i.e. $F \leq_L^2 G$ and $F \leq_{\bar{L}}^2 G$ do not imply the LD. Moreover, note that $F \leq_L^2 G$ implies that L_F starts above L_G and presents a lower (or equal) value of the Gini index. Differently, $F \leq_{\bar{L}}^2 G$ implies that L_F starts below L_G but still has a lower (or equal) Gini. Hence it is apparent that in both cases the condition $\Gamma(F) \leq \Gamma(G)$ is necessary to establish a dominance. In particular, if LCs cross only once, the one with higher Gini index dominates the other with respect to upward 2-LD (if the latter starts below, see [11]) or downward 2-LD (if the latter starts above, [6]).

As an alternative approach, Lando and Bertoli-Barsotti [6] combined the basic ideas (preferences) expressed by the $\leq_L^2$ and $\leq_{\bar{L}}^2$ orderings into a single preorder, which emphasizes inequality in both the tails of the distribution. This is done by symmetrically cumulating the Lorenz curve from both sides, through the disparity curve Δ_F, described above. In fact, it can be reasonable to wish that F is preferable to G if Δ_F is (uniformly) as small as possible, compared to Δ_F: this concept can be expressed as follows in terms of an integral inequality.

Definition 3. We say that F second-degree disparity dominates (2-DD) G and write $F \leq_D^2 G$ iff:

$$\int_0^t \Delta_F(p)dp \leq \int_0^t \Delta_G(p)dp \; \forall t \in [0,0.5].$$

Like $F \leq_L^2 G$, also $F \leq_D^2 G$ implies the condition $\int_0^{1/2} \Delta_F(p)dp \leq \int_0^{1/2} \Delta_G(p)dp$, which is equivalent to $\Gamma(F) \leq \Gamma(G)$ ($\int_0^{1/2} \Delta_F(p)dp = \int_0^1 (p - L_F(p))dp$).

Similarly to what has been discussed above, with regard to single-crossing LCs, if the curves are double-crossing, a theorem in [6] determines a sufficient condition for the 2-DD based on the positions of the crossing points and the values of the Gini index. Thus, 2-LD and 2-DD are especially suitable for ranking single-crossing and double-crossing LCs, respectively.

3 Empirical analysis

The main objective of this paper is to investigate the discriminating power of the dominance relations described in section 2. As a general rule, the stronger the preorder, the less it is applicable. For instance, we may find, in practice, many couples of intersecting LCs that are not comparable with LD. Thus, the main advantage of using a weaker criterion is that we can significantly increase the number of pairs of LCs that are comparable. Here we refer to the proportion of couples that are ranked according to a given preorder as the *rate of completeness* of such preorder. By increasing the rate of completeness we reduce the ambiguousness and we provide a normative justification for the use of an index of inequality that is coherent with the preorder considered, such as the Gini index. To perform our empirical analysis, we downloaded the income distributions of the EU-28 countries, plus some other non-EU counties, in the year 2015. The dataset is described as follows.

Data have been retrieved online from Eurostat's website. In particular, Eurostat provides the "distributions of income by quantiles" with two options in terms of income and living conditions indicator, namely: i) *top cut-off point*, which represent the income of the individual at the right end of the given quantile and; ii) *share of national equivalized income*, which is the share of the total income which belongs to a given interval. Eurostat provides i) and ii) for the three quartiles, the four quintiles, the nine deciles and the first (and last) five percentiles. Therefore, by properly cumulating the shares of national equivalized income we can obtain the values of the LC for

$p = 0.01,0.02,0.03,0.04,0.05,0.1,0.2,0.25,0.3,0.4,0.5,0.6,0.7,0.75,0.8,0.9,0.94,0.95,0.96,0.97,0.98,0.99,$

that is, a LC with 22 nodes.

For the year 2015, we computed the LCs of the EU-28 plus the following five non-EU countries: Macedonia, Iceland, Norway, Serbia and Switzerland, which totalize 33 countries. Then, we compared every pair of LCs (i.e. 33×32/2=528 pairs) based on the four dominance relations studied in section 2, namely LD, upward 2-LD, downward 2-LD and 2-DD. To be noted is that many countries present negative values for the LC, always for $p \leq 0.04$. The rates of completeness are shown in Table 1.

LD	Up. 2-LD	Down. 2-LD	2-DD
0.462	0.634	0.729	0.733

Table 1 Rates of completeness

This results reveal that the LD cannot even rank half of the pairs. This is due to the fact that the countries (especially the EU-28) present similar concentration patterns as well as similar values of the Gini index, thus LCs cross most of the times. Furthermore, we observe that generally the set of countries (at least as it is reported by Eurostat's data) concentrate most of their inequality in the right tails (i.e. high incomes) rather that in the left tails, so that cumulating from the right (downward 2-LD) is more critical (with regard to the discriminating power) than cumulating from the left (upward 2-LD). However, the relevance of the basic idea of downward 2-LD, that is, to emphasize the left tail, is acknowledged my many economists (unlike upward 2-LD). Therefore, it seems to be wiser to adopt 2-DD, so that we can rank the highest possible number of pairs by emphasizing both tails at the same time. However, although 2-DD yields the highest rate of completeness, it is far from reaching a complete ranking of the EU-28. In fact, some of the countries present very similar LCs, which might be ranked only by relying on dominance relations that are even weaker than 2-LD and 2-DD, so that their use would be hardly justifiable.

Further analyses, reported in Table 2, show the effectiveness of the different ranking criteria depending on the number of times a couple of LCs intersect. Downward 2-LD and 2-DD rank 65% of the single-crossing LCs, but 2-DD has slightly more discriminating power when dealing with double-crossing LCs, coherently with the theoretical result of [6]. We also note that the maximum number of times a LC crosses another is 3, but none of the dominance relations analysed is able to rank such pairs (that are just 3 pairs over 528, i.e. the 0.05%).

No. of intersections	Up. 2-LD	Down. 2-LD	2-DD
1	0.35	0.65	0.65
2	0.29	0.30	0.33

Table 2 Proportions of single and double-crossing LCs ranked by 2-LD and 2-DD

Table 3 shows the values of the different values of the Gini index for the EU-28 countries and the other five non-EU countries, and the different rankings yielded. We also computed the normalized Gini index Γ^*, which accounts for the presence of negative incomes, but in our case the differences between the normalized and non-normalized values are negligible (the rankings are identical), thus we do not report the values of Γ^* below (the maximum value of Γ_{max} obtained is 1.0001). However, our results presented above suggest that ranking based on the Gini index (when dealing with the 33 countries considered here) is supported by the LD only in less than 50% of the cases, whilst e.g. the 2-DD supports the Gini index in almost 75% of the cases.

Country	Γ	Country	Γ
Austria	0.541,{24}	**Lithuania**	0.754,{2}
Belgium	0.519,{26}	**Luxembourg**	0.564,{20}
Bulgaria	0.732,{4}	**Malta**	0.561,{21}
Croatia	0.610,{14}	**Macedonia**	0.672,{10}
Cyprus	0.668,{11}	**Netherlands**	0.530,{25}
Czech Rep.	0.502,{27}	**Norway**	0.479,{31}
Denmark	0.547,{23}	**Poland**	0.607,{15}
Estonia	0.689,{6}	**Portugal**	0.675,{9}
Finland	0.501,{28}	**Romania**	0.744,{3}
France	0.581,{19}	**Serbia**	0.759,{1}
Germany	0.597,{16}	**Slovakia**	0.470,{32}
Greece	0.681,{8}	**Slovenia**	0.488,{30}
Hungary	0.561,{22}	**Spain**	0.687,{7}
Iceland	0.469,{33}	**Sweden**	0.499,{29}
Ireland	0.592,{17}	**Switzerland**	0.586,{18}
Italy	0.642,{12}	**U.K.**	0.640,{13}
Latvia	0.705,{5}		

Table 3 Gini indices and rankings for the EU-28 countries and the other non-EU countries

It is also worth mentioning some special cases. The LC of Czech republic LD-dominates 19 of the other 32 countries. However, it can be seen in Figure 1 that the left tail of the LC of Czech republic is above all the others. Indeed, by using upward 2-LD, downward 2-LD and 2-DD, one obtains that Czech Republic dominates 24, 19 and 21 countries, respectively. The reason is that upward 2-LD emphasizes the left tail.

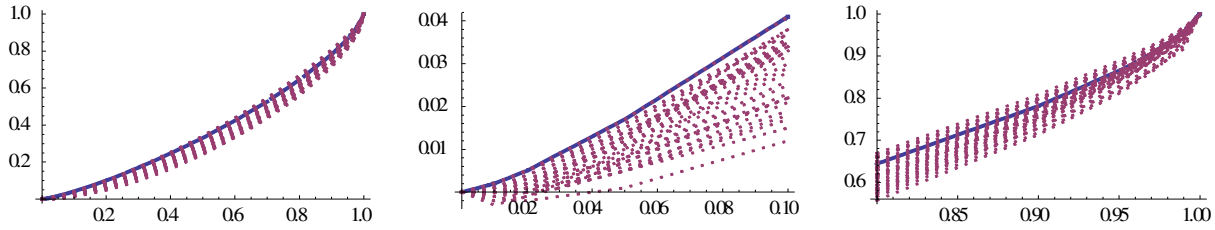


Figure 1 LCs of Czech Republic (solid form) and the other 32 countries (dotted form)

Another interesting case is represented by Iceland, which obtains the minimum value of the Gini index, although, strangely, it does not dominate any of the other countries because it presents also the minimum value of the LC for p=0.01 (see figure 2). Indeed, by using upward 2-LD, downward 2-LD and 2-DD, one obtains that Iceland dominates 0, 27 and 21 countries, respectively. The reason is that downward 2-LD) emphasizes the right tail.

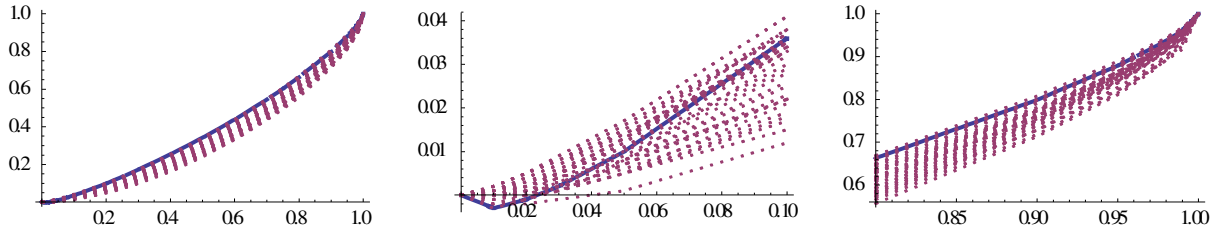


Figure 2 LCs of Iceland (solid form) and the other 32 countries (dotted form)

The examples above let us argue that one should find a compromise between the upward and downward 2-LD approach, in order to obtain more balanced results. This may be represented by the 2-DD, which emphasizes both tails at the same time.

4 Conclusion

Analysis of recent statistics on income inequalities in the European Union is undoubtedly a current topic. Comparisons of standards of living between countries are frequently based on gross domestic product (GDP) per capita, which presents in monetary terms how rich is a country compared to another. However, this headline indicator says very little about the distribution of income within a country and also fails to provide information in relation to non-monetary factors that may play a significant role in determining the quality of life of a particular population. On the one hand, inequalities in income distribution may create incentives for people to improve their situation through work, innovation or acquiring new skills. On the other hand, such income inequalities are often viewed as being linked to crime, poverty and social exclusion. Policymakers in general cannot combat poverty and social exclusion without analysing the inequalities within society, whether they are economic or social in nature.

In the paper, income inequality was measured by the Gini coefficient and represented by Lorenz curve. The shapes of the LCs of the EU-28 are somewhat similar and the main differences can be found in the right tails (high incomes) rather than in the left one. This also implies that downward 2-LD is effective, although most economists recommend ranking LCs on the basis of upward 2-LD. However, 2-DD combines upward and downward 2-LD so that it seems to be a wise compromise between the two approaches, besides providing the highest rate of completeness. The properly measured EU-wide Gini coefficient of disposable income inequality shows that inequality in the EU as whole are still real and not negligible. In Table 3, the Gini index is shown for the whole sample of evaluated countries. In the case of 33 European countries, the value of the Gini index in

2015 ranged from 0.469 (the best results for Iceland, Slovakia and Norway) to 0.759 (the worst results in Serbia followed by Lithuania and Romania). We observe that the most extreme values of the Gini index are obtained for non-EU countries (Serbia and Iceland). Countries at the top of the ranking are Slovenia, Sweden and Finland with the Gini indices from 0.488 to 0.501, whilst those at the bottom are Bulgaria, Latvia and Estonia, having the Gini index from 0.732 to 0.689. Then, the remainder countries can be broadly divided into two groups, with Croatia, Cyprus, Macedonia, Greece, Italy, Poland, Portugal and United Kingdom having the Gini index up to 0.6 value, and others such as Austria, Belgium, Czech Republic, Denmark, France, Germany, Hungary, Ireland, Luxembourg, Malta, Netherlands, Spain and Switzerland having values of the Gini index up to 0.5. Thus, we have two quite large sets with very similar values of the Gini index. Moreover, dominance results show that the Gini index is not sufficient to capture the real differences between the income distributions.

Within the EU, there are differences in income inequality which require policy action. It is important to remember that high levels of income inequality have various negative consequences (including poor health outcomes, weak social mobility and potential swings towards political populism). Therefore, efforts to address income inequalities should be stepped up in a number of countries as well as at the EU level. Many politicians and economists believe that economic growth replaces or diminishes the need for social policies. However, the EU growth over the last decades has been accompanied by increase in inequalities in many countries. Inequalities threaten social cohesion and they threaten growth. Inequality is a key problem facing the EU, and it has significant impacts not only on human well-being, but also on economic performance. If such concerns are correct, it is essential not only to build institutional structures for European social union but also to map social inequalities (including income dispersion) in the EU. Cohesion is about income and employment, but also about other dimensions of well-being.

The recent interest in income inequality is thus simply the recognition of the centrality of the topic to economic theory, policy, and performance. The recent return of the topic of income inequality has been triggered by important contributions to the empirical analysis of inequality, but these empirical analyses must be combined with an economic theory that is adequate to address the macroeconomic and microeconomic effects of income inequality on social welfare. These problems are not always well diagnosed because the empirical measurement of inequality is often unable to take into account the geographical dimension of inequality, which is particularly complex in the EU. To study income inequality in the EU as a whole, one needs adequate statistical tools which can be used in the geographical and political context faced by the EU. Important is also the relevant database as source for required data and indicators. In the case of the EU analysis, the relevant source is Eurostat, i.e. the official statistical office of the EU, but there are also some data particularities in Income and living conditions statistics, e.g. negative incomes. A deeper exploration of the underlying data shows that among those whose income is in the lower tail, negative income components are at work; these negative components can be due to self-employment, tax burden, transfers to other households or loss in property income.

Acknowledgements

This paper is supported by the grant No. 17-23411Y of the Czech Science Agency (to T.L and M.S.).

References

[1] Aaberge, R.: Ranking intersecting Lorenz curves. *Soc. Choice Welf* **33** (2009), 235-259.
[2] Atkinson, A.B.: More on the measurement of inequality. *J. Econ. Inequal.* **6** (2008), 277-283.
[3] Chen, C., Tien-Wang, T., and Tong-Shieng, R.: The Gini coefficient and negative income. *Oxford Economic Papers* **34** (3) (1982), 473-478.
[4] Davies, J.B. and Hoy, M.: Making inequality comparisons when Lorenz curves intersect. *Am. Econ. Rev.* **85** (1995), 980-986.
[5] Gastwirth, J.L.: A general definition of the Lorenz curve. *Econometrica* **39** (1971), 1037-1039.
[6] Lando, T. and Bertoli-Barsotti, L.: Weak orderings for intersecting Lorenz curves. *Metron* **74** (2), (2016), 177-192.
[7] Makdissi, P. and Yazbeck, M.: On the measurement of plutonomy. *Soc. Choice Welf* **44** (2015), 703-717.
[8] Marshall, A.W., Olkin, I. and Arnold, B.C.: *Inequalities: theory of majorization and its applications*. (second edition), Springer, 2009.
[9] Muliere P., Scarsini M.: A note on stochastic dominance and inequality measures. *Journal of Economic Theory* **49** (2) (1989), 314-323.
[10] Shorrocks, A.F., Foster J.E.: Transfer sensitive inequality measures. *Rev. Econ. Stud.*, **14** (1987) 485-497.
[11] Zoli C.: Intersecting generalized Lorenz curves and the Gini index. *Soc. Choice Welf* **16** (2), (1999) 183-196.

On modelling the evolution of financial metrics in decision making unit in the electronics industries

Eva Litavcová[1], Sylvia Jenčová[2], Petra Vašaničová[3]

Abstract. The electrical industry is one of the largest industries in the world. In Slovakia, it forms, together with the mechanical engineering sector, the main pillar of industry for the country. It is growing at the fastest pace among all manufacturing sectors, and consequently represents one of the most attractive sectors for foreign investors. In this paper, we focus on the analysis of the electrical industry in Slovakia, concentrating on one of the most successful, Tesla Stropkov, Inc., using the NACE 263 - Manufacture of Communication Equipment standards. Other analyzes show Tesla Stropkov to be within the 60 most successful enterprises of electrical industry in the Slovak Republic. Our analysis tries to determine if we can expect future positive development of operations for the company being studied. Variables of production, returns, costs and revenues of the selected company were analyzed monthly from 2009 to 2016. This paper's modelling assumptions concerning the future performance of these variables for the selected company using exponential smoothing and Box-Jenkins methodology. It shows stable development of volume of production and sales, with revenues forecasts showing steady slightly decreasing future development, and vice versa, at the cost of moderate growth.

Keywords: Electrical industry, Financial metrics, Time series.

JEL Classification: C22
AMS Classification: 91B84

1 Introduction

In Slovak Republic, electrical industry is one of the largest employers in industry, second of the most important exporters (after engineering). Driving factor of development of electrical industry which is the third biggest industry (year 2015) is automotive industry. Electrical industry has a quarter share on industrial export of this country, where electrical company Samsung Electronics Slovakia was the largest exporter in Slovakia in 2009. According to cumulative receipts of electrical companies was this industry on the second place after engineering. Nowadays, every fifth TV in European Union is made in Slovakia. The importance of this industry is quantified by the value added per employee. This industry is the more important contributor to creating value added in manufacturing (11.56%) and is the more important employer (12.54% of total employment, average number of employees was 43.039). This industry had the revenues 9.279 mil. Euro, export 8.215 mil. Euro, net income 257 mil. Euro in 2015. The competitiveness of electrical industry is determined by ability of its decision-making units or ability of industry reacts on the challenges by its competitors. According to another work of this paper authors (in press), one of the successful companies in the electrical industry in Slovakia is Tesla Stropkov, Inc. with it's 38th position in the rankings of success, based on coefficient of competitiveness, among the 60 most successful electrical firms. According to volume of revenues in 2015, this company held 43rd place in the non-financial corporation of the electrical industry. By using the application of coefficients of competitiveness [2], the company was included in Group A - competitive companies. In use of multivariate comparative method of distance from the fictitious object [16], this company was ranked 22nd. This method points to the distance of company from the ideal object in all of the indicators (basic production power of the company, return on sales, financial performance per one employee, and financial labour productivity). According to method of standardized variable, it occupied the 19th place within the electrical industry.

Tesla Stropkov, Inc. is an important employer in the region, in which is one of the highest unemployment rate in Slovakia. Therefore, it is an interesting question whether the future positive development of this company

[1] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, eva.litavcova@unipo.sk.
[2] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, sylvia.jencova@unipo.sk.
[3] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, petra.vasanicova@smail.unipo.sk.

can still be expected. Strategic business performance management in Slovak industry was studied in [15]. The aim of this paper is to model the development of selected financial metrics of Tesla Stropkov, Inc. Because our analysis requires own internal company data, this analysis may be applicable to forecasting and valuation projects for a variety of manufacturing companies. For modelling and forecasting financial development it this study we use two models: the exponential smoothing model, and the Box - Jenkins Seasonal Autoregressive Integrated Moving Average (SARIMA) model. Mentioned methodology was used in number of studies related to forecast of financial metrics, in [14] to forecast of consumer retail sales, in [11] to predict long-term earnings, in [13] to model revenues from services, for example.

2 Data

To model production, returns, costs, and revenues of the selected company are used corresponding monthly unadjusted internal company data of these variables in the period from January 2009 to October 2016. Dataset was obtained from financial manager of the company.

3 Methodology

To purpose the objective of this paper, we model the current and future development of selected financial metrics in company Tesla Stropkov, Inc. by using the exponential smoothing model, and the Box - Jenkins Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The relevant literature dealing with statistical forecasting extends to the development of the simple exponential smoothing method by Brown [6] in the 1950's and some of its extensions by Holt [9] and Winters [18]. In the1970s, ARIMA models were developed and have been studied extensively by many researchers. Their theoretical underpinnings were described by Box and Jenkins [4] and later by Box et al. [5]. Nowadays are also widely used more complex models and associated statistical tests proposed, for example, by Engle [7], Bolerslev [3], Granger [8] and so on, as clearly stated for example in [1], [19].

3.1 Exponential smoothing

Using exponential smoothing to forecasting time series is the most common. In this paper, we use simple seasonal, Holt-Winters additive and Holt-Winters multiplicative methods. Forecast equation of the component form for the additive method [18], [12] is:

$$F_{t+m} = L_t + b_t m + S_{t-s+m}, \tag{1}$$

where F_{t+m} is the forecast for m periods ahead, s is the length of seasonality. Method estimates three components, L_t is the level of the series, b_t is the trend, S_t is the seasonal component in time t. Smoothing equations for these three components are:

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}), \tag{2}$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}, \tag{3}$$

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s}, \tag{4}$$

where Y_t is the observed values. Appropriate smoothing constants α, β, γ, which can be obtain from the range from 0 to 1 are determined subjectively by users on the basis of their own experience. The component form of Holt - Winters' multiplicative seasonal model [10], [12], [18] is:

$$F_{t+m} = (L_t + b_t m)S_{t-s+m}, \tag{5}$$

Trend component is the same equation like (3), level and seasonal component can be written as:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}), \tag{6}$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s}, \tag{7}$$

The error correction representations of Holt-Winters additive and Holt-Winters multiplicative methods and starting conditions are not reported there.

3.2 Box-Jenkins SARIMA model

According to [4], [5], [17], a simple equation to define the autoregressive moving average (ARMA)(p, q) model for a stationary time series is given below:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} . \quad (8)$$

The first term in ARIMA model represents an autoregressive (AR) term of the order *p* having the form of

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t . \quad (9)$$

This (AR) term refers to the current time series values Y_t as a function of past time series values Y_{t-1}, Y_{t-2},..., Y_{t-p}. The ϕ_1, ϕ_2, ϕ_3 are autoregressive coefficients that relates Y_t to Y_{t-1}, Y_{t-2},..., Y_{t-p}.

The moving average MA(q) term of the model is represented as,

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} \quad (10)$$

where, $\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-q}$ are the past random shocks or independent white noise sequence with mean = 0 and variance = σ^2; $\theta_1, \theta_2, ..., \theta_q$ are the moving average coefficients relating Y_t to $\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-q}$.

When the (AR) and (MA) specifications are combined together with integration (differencing) term, they constitute an ARIMA (p,d,q) model, where p, d and q indicate orders of autoregression, differencing and moving average. The model is mathematically given as

$$(1-B)^d Y_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t , \quad (11)$$

where, *t* denotes the time indices, *B* is the backshift operator, ie., $BY_t = Y_{t-1}$. $\phi(B)$ and $\theta(B)$ are the autoregressive and moving average operators respectively and can be written as

$$\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - ... - \phi_p B^p \quad (12)$$

$$\theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - ... - \theta_q B^q . \quad (13)$$

Seasonality is a pattern which is repeating itself over a fixed time interval. Here, the monthly dataset is presenting a seasonal period of 12 months. In order to obtain a stationarity, seasonal differencing is performed by taking difference between the present and corresponding observation from the previous year. Taking into consideration the seasonality of our time series, a seasonal ARIMA denoted as SARIMA (p, d, q) x $(P, D, Q)_s$, is used, where P,D,Q represent seasonal autoregressive, differencing and moving average orders respectively and *s* is number of seasons. For the present study, *s* = 12. SARIMA(p, d, q)$(P, D, Q)_s$ built for the time series is defined as:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t , \quad (14)$$

where, *B* is the backshift or lag operator, *s* is the seasonal lag (in 'month' for present study); ε_t represents error variables; *d* and *D* are non-seasonal and seasonal differences; ϕ and Φ are the non-seasonal and seasonal autoregressive parameters; θ and Θ are the non-seasonal and seasonal moving average parameters; $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ are seasonal parts of (AR) and (MA) specifications respectively.

3.3 Goodness-of-Fit Statistics

For determination of validity of our model, we use criterion of the mean absolute percentage error (MAPE). According to [10], MAPE is given by:

$$MAPE = \sum_{i=1}^{n} \left| \frac{(Y_i - F_i)/Y_i}{n} \right| \cdot 100 . \quad (15)$$

In this equation X_i is the actual data for period *i*, F_i denotes the forecast for period *i*, and *n* is the number of observations. MAPE ≤ 10% means highly accurate forecast, MAPE = 10 – 20% means good forecast, MAPE =

20 – 50% denotes reasonable forecast, and MAPE > 50% denotes inaccurate forecasting. Another used criterion is stationary R-squared, by [20] is given by:

$$R_S^2 = 1 - \frac{\sum_t (Y_t - \hat{Y}_t)^2}{\sum_t (\Delta Y_t - \overline{\Delta Y})^2}. \tag{16}$$

Stationary and usual R-squared can be negative with range $(-\infty, 1]$. A negative R-squared value means that the model under consideration is worse than the baseline model. Zero R-squared means that the model under consideration is as good or bad as the baseline model. Positive R-squared means that the model under consideration is better than the baseline model.

4 Results

Using exponential smoothing, the best found models for four selected variables are presented in the Table 1. To test null hypothesis about no autocorrelation in residuals Ljung-Box statistics with 15 d.f. has been used. In all four models this null hypothesis was not rejected. Normality of residuals has been checked in all models and it has been shown residuals from all four models are i.i.d.. Outliers has been not detected in all models, too.

Variable	Model	Level	Trend	Season	MAPE	Stationary R^2
Production	Winters' Additive	0.201[c]	0.000	0.000	8.368	0.642
Returns	Winters' Additive	0.228[c]	0.209[b]	0.001	7.091	0.639
Costs	Winters' Multiplicative	0.076[b]	0.577[a]	0.152[a]	7.408	0.751
Revenues	Simple Seasonal	0.400[d]	–	0.000	8.737	0.558

Table 1 Exponential smoothing parameters of financial variables

According to MAPE and Stationary R^2 interpretation, we can say, that our models represent highly accurate forecast. Forecasted values with lower (LCL) and upper (UCL) 95% confidence limits, obtained by exponential smoothing of production, returns, costs, and revenues of the selected company are presented in Table 2. Observed, smoothed and forecast values of exponential smoothing models can be seen in Figure 1.

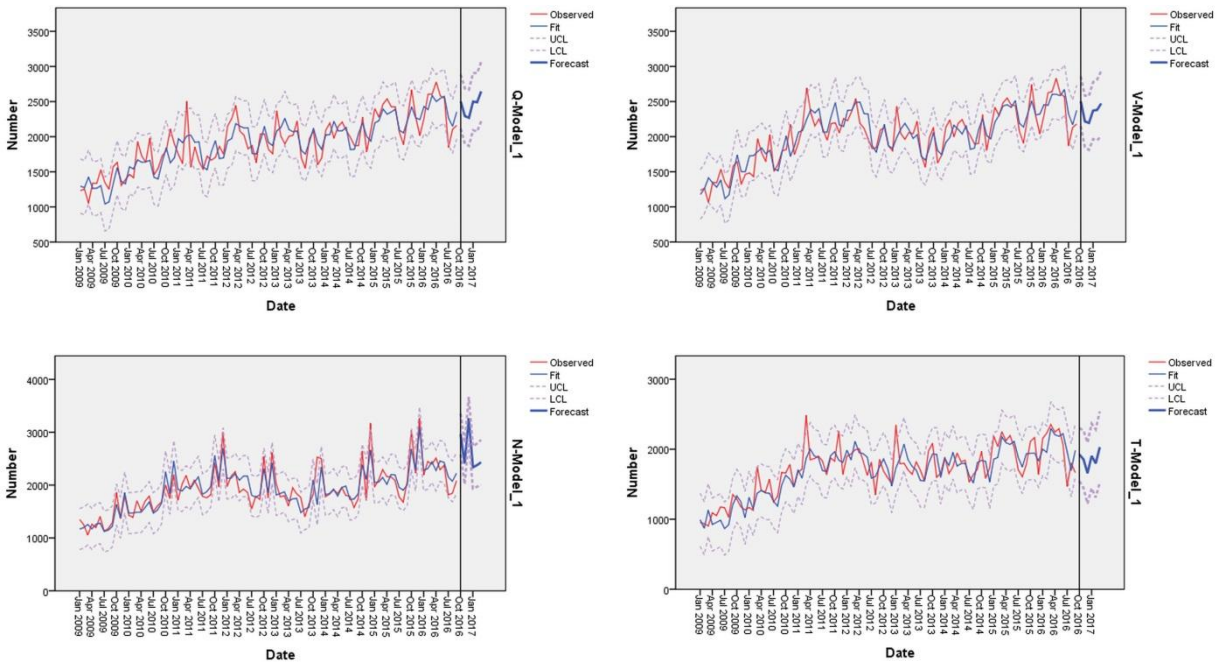


Figure 1 Observed, smoothed and forecast series with 95% confidence bounds of production, returns, costs and revenues using exponential smoothing models

		Oct 2016	Nov 2016	Dec 2016	Jan 2017	Feb 2017	Mar 2017
Q	LCL	2115	1900	1868	2092	2075	2223
Production	Forecast	2501	2293	2268	2500	2490	2646
	UCL	2886	2686	2669	2908	2905	3068
V	LCL	2150	1850	1811	1965	1949	2010
Returns	Forecast	2504	2217	2195	2371	2381	2473
	UCL	2857	2583	2580	2777	2813	2935
N	LCL	2577	2018	2868	1935	1967	2008
Costs	Forecast	2965	2408	3267	2335	2376	2433
	UCL	3353	2798	3666	2734	2786	2857
T	LCL	1545	1449	1224	1443	1312	1522
Revenues	Forecast	1924	1858	1659	1904	1798	2031
	UCL	2304	2266	2095	2366	2284	2540

Table 2 Forecast values of financial variables using exponential smoothing models

In comparing with results of exponential smoothing, seasonal ARIMA with seasonal lag 12 has been used. Taking into account assumptions and data fitting, we have chosen the best parsimony models (to test residuals were used Ljung-Box statistics and K-S test of normality, to compare models AIC and BIC were they used, too). The resulting models with parameters and goodness-of-fit statistics are presented in Table 3. All models are without constants.

Variable	model	θ_1	Φ_1	Θ_1	MAPE	Stationary R^2
Production	ARIMA(0,1,1)(1,0,0)	0.752[d]	0.498[d]	–	9.135	0.406
Returns	ARIMA(0,1,1)(0,1,1)	0.664[d]	–	0.652[d]	8.359	0.435
Costs	ARIMA(0,1,1)(0,1,1)	0.748[d]	–	0.736[d]	8.787	0.648
Revenues	ARIMA(0,1,1)(1,0,0)	0.685[d]	0.503[d]	–	8.856	0.448

Table 3 SARIMA models of financial variables

Forecast values with lower (LCL) and upper (UCL) 95% confidence limits, obtained by SARIMA models of production, returns, costs, and revenues of the selected company are presented in Table 4.

		Oct 2016	Nov 2016	Dec 2016	Jan 2017	Feb 2017	Mar 2017
Q	LCL	2050	1840	1699	1808	1966	1955
Production	Forecast	2491	2295	2167	2288	2459	2460
	UCL	2933	2750	2634	2768	2952	2964
V	LCL	2154	1747	1716	1992	2055	2087
Returns	Forecast	2574	2190	2181	2478	2561	2612
	UCL	2993	2632	2645	2963	3066	3136
N	LCL	2359	1849	2681	1791	1884	1938
Costs	Forecast	2822	2327	3173	2297	2403	2470
	UCL	3285	2805	3665	2802	2921	3001
T	LCL	1554	1561	1275	1516	1537	1581
Revenues	Forecast	1970	1997	1730	1990	2029	2090
	UCL	2386	2432	2185	2464	2520	2598

Table 4 Forecast values of financial variables using SARIMA models

Resulting SARIMA models for production, returns, costs and revenues can be gradually written as:

$$Y_t = Y_{t-1} + 0.498Y_{t-12} - 0.498Y_{t-13} + \varepsilon_t - 0.752\varepsilon_{t-1}, \tag{17}$$

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + \varepsilon_t - 0.664\varepsilon_{t-1} - 0.652\varepsilon_{t-12} + 0.433\varepsilon_{t-13}, \tag{18}$$

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + \varepsilon_t - 0.748\varepsilon_{t-1} - 0.736\varepsilon_{t-12} + 0.551\varepsilon_{t-13}, \tag{19}$$

$$Y_t = Y_{t-1} + 0.503Y_{t-12} - 0.503Y_{t-13} + \varepsilon_t - 0.685\varepsilon_{t-1}. \tag{20}$$

Even though both models represents high accurate forecast, goodness-of-fit statistics values are more acceptable for exponential smoothing models. In comparing results of forecasts, models are very similar. Forecast values for variable *Production* are estimated higher by exponential smoothing model than by SARIMA and vice versa, for variable *Revenues* are all forecasted values from SARIMA model higher than from exponential smoothing. For variables *Returns* and *Costs*, estimates vary differently.

5 Conclusion

In this paper was presented modelling of selected financial metrics of company Tesla Stropkov, Inc. using the exponential smoothing models and the Box - Jenkins SARIMA models. Contribution of this work can be useful for investigated company and its region. We can expect that development of Tesla Stropkov, Inc. is stabilised and the employment of the region will not be reduced by this company.

Acknowledgements

Supported by the grant No. 1/0945/17 of the Grant Agency VEGA.

References

[1] Arlt, J. and Arltová, M.: *Ekonomické časové řady*. Professional Publishing, Praha, 2009.
[2] Chajdiak, J.:. Konštrukcia vybraných mier konkurencieschopnosti. In Forum Statisticum Slovacum, **11**(5) (2016), 16-21.
[3] Bollerslev, T.: Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31** (1986), 307-327.
[4] Box, G. E. P., Jenkins, G. M.: *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 1970.
[5] Box, G. E. P., Jenkins, G. M., and Reinsel, G.C.: *Time Series Analysis: Forecasting and Control*. Prentice-Hall, Englewood Cliff, N.J., 1994.
[6] Brown, R. G.: *Statistical Forecasting for Inventory Control*. McGraw-Hill, New York, 1959.
[7] Engle, R. F.: Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50** (1982), 987–1007.
[8] Granger, C. W. J.: Time Series Analysis, Cointegration, and Applications. *American Economic Review*, **94** (3) (2004), 421–425.
[9] Holt, C. C.: *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*. O.N.R. Memorandum no. 52. Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1957.
[10] Lee, C. K., Song, H. J., and Mjelde, J. W.: The forecasting of International Expo tourism using quantitative and qualitative techniques. *Tourism Management* **29** (2008), 1084-1098.
[11] Lorek, K. S., and Willinger, G. L.: An analysis of the accuracy of long-term earnings predictions. *Advances in Accounting* **19** (2002), 161-175.
[12] Makridakis, S., Wheelwright, S., and Hyndman, R. J.: *Forecasting: methods and applications*. John Wiley & Sons, USA, 1998.
[13] Porto, R. B.: The commercial cycle from the viewpoint of operant behavioral economics: effects of price discounts on revenues received from services. *Revista de Administração* **51**(3) (2016), 310-322.
[14] Ramos, P., Santos, N., and Rebelo, R.: Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and computer-integrated manufacturing* **34** (2015), 151-163.
[15] Rajnoha, R., Stefko, R., Merková, M., and Dobrovic, J.: Business intelligence as a key information and knowledge tool for strategic business performance management. *E+M Ekonomie a Management*, **19**(1) (2016), 181-203.
[16] Stankovičová, I., and Vojtková, M.: Viacrozmerné štatistické metódy s aplikáciami. Bratislava, Iura Edition, 2007.
[17] Taneja, K., Ahmad, S., Ahmad, K., and Attri, S. D.: Time series analysis of aerosol optical depth over New Delhi using Box–Jenkins ARIMA modeling approach. *Atmospheric Pollution Research* **7** (2016), 585-596.
[18] Winters, P.: Forecasting sales by exponentially weighted moving averages. *Management Science* **6** (1960), 324–342.
[19] Wooldridge, J. M.: *Introductory Econometrics*. A Modern Approach. Cengage learning, Boston, 2016.
[20] IBM SPSS help: Available at: https://www.ibm.com/support/knowledgecenter/SSLVMB_21.0.0/com.ibm.spss.statistics.help/alg_tsmodel_gof_.htm

Correlation Structure of Underlying Assets Affecting Multi-asset European Option Price

Ladislav Lukáš[1]

Abstract. The paper deals with formulation of multi-asset European put option pricing problems assuming spot prices of underlying asset to follow geometric Brownian motion with correlation structure. Using traditional approach based on self-financing portfolio and application of Itô's formula to option price, we get resulting partial differential equation describing the evolution of option price in space and time. The payoff function can take various form, two basic ones are presented in particular. Further, we discuss general concept which generates a set of low-dimensional initial-boundary value problems after time reversing substitution has been applied to original terminal-value problem. Numerical solution is focused on two-dimensional basket European put option problem. First, the problem is recast to variational formulation providing convenient platform for application of finite element method. Next, we present results of several numerical experiments to analyze influence of correlation structure of underlying assets upon option price. All calculations are performed by open source sw package FreeFem++.

Keywords: multi-asset European option price, basket put option, variational formulation, finite element method.

JEL classification: G13
AMS classification: 91G80

1 Introduction

First, we make a short overview of multi-asset European put option pricing problem formulation assuming spot prices of underlying asset follow geometric Brownian motion with correlation structure. Using traditional approach based on self-financing portfolio and application of Itô's formula to option price, one gets resulting partial differential equation describing the evolution of option price in space and time. Such approach is discussed in [1], [5], and [7], with different level of details and abstraction.

We consider n risky assets A_i with their spot prices $S_{i,t}, i = 1, \ldots, n$, at time t, where t counts the time elapsed since the entry into an option contract. We assume the $S_{i,t}$ to satisfy the following stochastic differential equations

$$\mathrm{d}S_{i,t} = S_{i,t}(\mu_i \mathrm{d}t + \sigma_i \mathrm{d}B_{i,t}), \quad i = 1, \ldots, n,\ t \geq 0. \tag{1}$$

where μ_i are expected rate of returns (constants), σ_i are volatilities (positive constants), and $\{B_{i,t}\}, i = 1, \ldots, n$ are possibly correlated Brownian motions, with matrix of correlation factors $\rho_{i,j},\ -1 \leq \rho_{i,j} \leq 1,\ i, j = 1, \ldots, n$, where $\rho_{i,j}$ is correlation factor (constant) between $\{B_{i,t}\}$ and $\{B_{j,t}\}$, thus giving the covariance $\mathrm{Cov}(S_{i,t}, S_{j,t}) = \rho_{i,j}\sigma_1\sigma_2$, between $\{S_{i,t}\}$ and $\{S_{j,t}\}$, as well

$$\mathrm{E}(\mathrm{d}B_{i,t}) = 0, \quad \mathrm{Var}(\mathrm{d}B_{i,t}) = \mathrm{d}t, \quad \mathrm{Cov}(\mathrm{d}B_{i,t}, \mathrm{d}B_{j,t}) = \rho_{i,j}\mathrm{d}t.$$

For pricing purposes, one assumes complete market formed from n risky underlying assets $A_i, i = 1, \ldots, n$, and one standard risk-free asset A^0, with its price S^0 governed by exponential deterministic process with risk-free rate $r > 0$, thus enabling construction of self-financing portfolio, in general.

Let $V(S_1, \ldots, S_n, t)$ be a basket option price compounded from S_i, at time t, where we write simply S_i instead of $S_{i,t}$ dropping a sub-index t, as usual. For classical formulation, one assumes the function V to be smooth in t, and twice differentiable by each S_i, just to admit application of Itô's formula thereon. The self-financing portfolio Π is constructed by Δ-hedging principle, again, using Δ_i shares of assets A_i

$$\Pi = V(S_1, \ldots, S_n, t) - \sum_{i=1}^{n} \Delta_i S_i. \tag{2}$$

[1]University of West Bohemia, Faculty of Economics, Dept. of Economics and Quantitative Methods, Univerzitní 22, 306 14 Pilsen, lukasl@kem.zcu.cz

Consequently, a self-financed dynamics of this portfolio stems from balancing two expressions for portfolio value increment $\mathrm{d}\Pi$ to be risk-free in time period $(t, t+\mathrm{d}t)$, assuming European option framework, as follows

$$\begin{aligned} \mathrm{d}\Pi &= \mathrm{d}V - \sum_{i=1}^{n} \Delta_i \mathrm{d}S_i, \\ \mathrm{d}\Pi &= r\Pi \mathrm{d}t = r(V - \sum_{i=1}^{n} \Delta_i S_i)\mathrm{d}t. \end{aligned} \tag{3}$$

where the first one gives a portfolio value increment due to intended changes of asset shares Δ_i, whilst the second one expresses increment of risk-free bank accout value during $\mathrm{d}t$.

Application of Itô's formula upon $\mathrm{d}V$, Δ-hedging principle yielding $\Delta_i = \frac{\partial V}{\partial S_i}$, $i = 1, \ldots, n$, and the arbitrage-free argument, leads finally to the Black-Scholes partial differential equation (B-S PDE) for n-D basket European option price V, written in following form

$$\frac{\partial V}{\partial t} + \mathcal{A}_n V - rV = 0, \quad \forall (S_1, \ldots, S_n, t) \in \mathbb{R}_+ \times \cdots \times \mathbb{R}_+ \times [0, T), \tag{4}$$

where the differential operator $\mathcal{A}_n$, also called infinitesimal generator of the Markov family of n state variables, i.e. $S_i \in \mathbb{R}_+$, $i = 1, \ldots, n$, is given by its application on any function $v(S_1, \ldots, S_n)$ of $n+1$ variables from the class $C^{2,1}(\mathbb{R}_+^n \times [0, T))$

$$\mathcal{A}_n v = r \sum_{i=1}^{n} S_i \frac{\partial v}{\partial S_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{i,j} \sigma_i \sigma_j S_i S_j \frac{\partial^2 v}{\partial S_i \partial S_j}. \tag{5}$$

For more technical details, we refer to [1], in particular.

The B-S PDE (4) holds for both European type call and put option pricing problems. However, within the European call option type framework, one needs additional assumption relating growth bounds for $|S_i \frac{\partial V}{\partial S_i}|$.

Denoting the n-D basket European put option payoff function at maturity, i.e. $t = T$, by $h_P(S_1, \ldots, S_n; K)$, with given strike price K, then the n-D basket European put option pricing problem, in classical formulation, as terminal value problem (TVP) for parabolic PDE with $(n+1)$ variables, is following

$$\begin{aligned} &\frac{\partial V}{\partial t} + \mathcal{A}_n V - rV = 0, \quad \forall (S_1, \ldots, S_n, t) \in \mathbb{R}_+^n \times [0, T), \\ &V(S_1, \ldots, S_n, T) = h_P(S_1, \ldots, S_n; K), \ \forall (S_i, \ldots, S_n) \in \mathbb{R}_+^n. \end{aligned} \tag{6}$$

The payoff function $h_P(S_1, \ldots, S_n; K)$ can be quite complicated. The payoff function can take various forms, two basic ones are presented in [2], Chapter 1, in particular. Among the simplest ones, we can list

- put option on a weighted sum:

$$h_P(S_1, \ldots, S_n; K) = \max(K - \sum_{i=1}^{n} \alpha_i S_i, 0), \ \sum_{i=1}^{n} \alpha_i = 1, \ \alpha_i \geq 0, \ \exists i, \alpha_i > 0,$$

- best-of put option:

$$h_P(S_1, \ldots, S_n; K) = \max(K - \max_{i=1,\ldots,n}(S_i), 0)$$

.

In case of n-D basket option and unlike to the option pricing problem depending upon one underlying asset only, we need to discuss also boundary conditions of function $V(S_1, \ldots, S_n, t)$ considering two basic cases

- case 1: $S_i \to +\infty$,
- case 2: at least for one S_i holds, $S_i = 0$.

The first case, $S_i \to +\infty$, may simply take over the well-known behaviour of 1-D option price: $\lim_{S \to +\infty} V(S, t) = 0$, $\forall t \in [0, T)$, thus giving

$$\lim_{S_i \to +\infty} V(S_1, \ldots, S_n, t) = 0, \quad i = 1, \ldots, n, \ \ \forall t \in [0, T). \tag{7}$$

The second case, at least for one S_i holds, $S_i = 0$, is more complicated, because there can vanish not only one, but two, three, or even $n-1$ asset prices S_i simultaneously. General concept, which generates a set of low-dimensional initial-boundary value problems after time reversing substitution has been applied to original terminal value problem, consists of precise hierarchical decomposition of all possible degeneration combinations of original n-D PDE.

2 2-D basket European put option pricing problem – numerical example

In classical formulation, assuming $V(S_1, S_2, t) \in C^{2,2,1}(\mathbb{R}^2_+ \times [0, T)$, and parameters $K, T, r, \sigma_1, \sigma_2, \rho$, are given, the price $V(S_1, S_2, t)$ has to satify

1. PDE on domain $\Omega = \mathbb{R}^2_+ \times [0, T)$

$$\begin{aligned} \frac{\partial V}{\partial t} + \frac{1}{2}(\sigma_1^2 S_1^2 \frac{\partial^2 V}{\partial S_1^2} + \sigma_2^2 S_2^2 \frac{\partial^2 V}{\partial S_2^2}) + \rho\sigma_1\sigma_2 S_1 S_2 \frac{\partial^2 V}{\partial S_1 \partial S_2} \\ + r(S_1 \frac{\partial V}{\partial S_1} + S_2 \frac{\partial V}{\partial S_2} - V) = 0, \quad \forall (S_1, S_2, t) \in \Omega, \end{aligned} \tag{8}$$

2. Terminal condition – case: the best-of put option payoff function

$$V(S_1, S_2, T) = \max(K - \max_{j=1,2}(S_j), 0), \quad \forall (S_1, S_2) \in \mathbb{R}^2_+, \tag{9}$$

3. Boundary conditions
   i) case: $S_i \to +\infty$, $i = 1, 2$:

$$\lim_{S_i \to +\infty} V(S_1, S_2, t) = 0, \quad S_j \in \mathbb{R}_+, \ (i, j) = \{(1, 2), (2, 1)\}, \quad \forall t \in [0, T) \tag{10}$$

   ii) case: $V : (S_1 \in \mathbb{R}_+, S_2 = 0, t) \to V(S_1, 0, t)$, $\forall t \in [0, T)$:

$$\begin{aligned} \frac{\partial V}{\partial t} + \frac{1}{2}\sigma_1^2 S_1^2 \frac{\partial^2 V}{\partial S_1^2} + r S_1 \frac{\partial V}{\partial S_1} - rV) = 0 \\ V(S_1, 0, T) = \max(K - S_1, 0), \quad \forall S_1 \in \mathbb{R}_+ \end{aligned} \tag{11}$$

   iii) case: $V : (S_1 = 0, S_2 \in \mathbb{R}_+, t) \to V(0, S_2, t)$, $\forall t \in [0, T)$:

$$\begin{aligned} \frac{\partial V}{\partial t} + \frac{1}{2}\sigma_2^2 S_2^2 \frac{\partial^2 V}{\partial S_2^2} + r S_2 \frac{\partial V}{\partial S_2} - rV) = 0 \\ V(0, S_2, T) = \max(K - S_2, 0), \quad \forall S_2 \in \mathbb{R}_+ \end{aligned} \tag{12}$$

where σ_1, σ_2, (constants), are volatilities of underlying assets S_1, S_2, and ρ, (const), is correlation factor between S_1, and S_2, $-1 < \rho < 1$, thus giving covariance $\mathrm{Cov}(S_1, S_2) = \rho\sigma_1\sigma_2$, between S_1 and S_2.

As the low-dimensional TVP-s (11) and (12) correspond to well-known European put option pricing problem, we set the input data in order to go along with the data for 2-D basket European put option pricing problem investigated by FEM later. Using the classical pricing formula, we get the results, i.e. prices $V(S, 0)$, which are depicted in Figure 4 (*left*), and listed in Table 1, for some selected values of underlying asset price S. Following

S_0	10^{-3}	10	20	30	40	50	60
$V_0(\sigma = .3)$	39.01	29.01	19.01	9.39	2.87	0.557	0.080
$V_0(\sigma = 1.2)$	39.01	29.37	21.78	16.35	12.49	9.713	7.668

Table 1: European put option price $V_0(S) = V(S, 0)$ with different σ.

[1], we are able to derive a weak formulation of 2-D basket European put pricing problem by carrying out four almost standard basic steps, which are also discussed in [6]

1. recast TVP into initial value problem (IVP) by time substitution $\tau = T - t$, thus giving $\partial/\partial t = -\partial/\partial\tau$,
2. select proper function space suitable for technical manipulation with integrals being endowed with convenient norm and scalar product,
3. select a test function, multiply the PDE by test function, and perform integration by parts,
4. formulate the IVP in weak sense, i.e. in variational formulation.

Application of $\tau = T - t$, and $\partial/\partial t = -\partial/\partial\tau$, cause changes, and PDE (8) is converted into (13)

$$V(S_1, S_2, t) \rightsquigarrow u(S_1, S_2, \tau), \quad V(S_1, S_2, T) \rightsquigarrow u(S_1, S_2, 0) = u_0(S_1, S_2),$$

$$\frac{\partial V}{\partial \tau} - \mathcal{A}_2(u(S_1, S_2, \tau)) + ru = 0, \quad \forall (S_1, S_2, \tau) \in \mathbb{R}^2_+ \times (0, T], \tag{13}$$

$$\mathcal{A}_2(u(S_1, S_2, \tau)) = \frac{1}{2}(\sigma_1^2 S_1^2 \frac{\partial^2 V}{\partial S_1^2} + \sigma_2^2 S_2^2 \frac{\partial^2 V}{\partial S_2^2}) + \rho\sigma_1\sigma_2 S_1 S_2 \frac{\partial V^2}{\partial S_1 \partial S_2} + r(S_1 \frac{\partial V}{\partial S_1} + S_2 \frac{\partial V}{\partial S_2}). \tag{14}$$

In general, variational equation of the IVP for parabolic PDE is expressed globally on whole time period $(0,T]$. However, for computational purposes, there is more useful a local version. As given in [1], the local variational equation takes the following form, where W is a suitable space of test functions v, and τ is replaced by t, as usual

$$\Big(\frac{\partial u(S_1,S_2,t)}{\partial t},v\Big)+a_t(u(t),v)=0,\ \forall v\in W,\ \text{a.e.}\, t\in(0,T), \tag{15}$$

$$a_t(u,v)=b_{2,t}(u,v)+b_{1,t}(u,v),\ \forall v\in W, \tag{16}$$

$$b_{2,t}(u,v)=\frac{1}{2}\int_Q\Big(\sigma_1^2S_1^2\frac{\partial u}{\partial S_1}\frac{\partial v}{\partial S_1}+\sigma_2^2S_2^2\frac{\partial u}{\partial S_2}\frac{\partial v}{\partial S_2}+\rho\sigma_1\sigma_2S_1S_2(\frac{\partial u}{\partial S_1}\frac{\partial v}{\partial S_2}+\frac{\partial u}{\partial S_2}\frac{\partial v}{\partial S_1})\Big)\mathrm{d}\Omega, \tag{17}$$

$$b_{1,\tau}(u,v)=r\int_\Omega\big(-S_1\frac{\partial u}{\partial S_1}-S_2\frac{\partial u}{\partial S_2}+u(S_1,S_2,t)\big)v\mathrm{d}\Omega, \tag{18}$$

$$W=\{v\,|\,v\in L^2(\mathbb{R}^2_+),\ S_i\frac{\partial v}{\partial S_i}\in L^2(\mathbb{R}^2_+),\ i=1,2\}. \tag{19}$$

Discretization of the problem and FEM solution, is given in [1], and [5], mainly. First, one has to make a *localization* of the problem, i.e. to truncate the unbounded domain $\Omega=\mathbb{R}^2_+$ of prices (S_1,S_2) into a bounded domain $G=(0,\bar S_1)\times(0,\bar S_2)$. Usually, one sets $\bar S_1=\bar S_2=\bar S$, which yields $G=(0,\bar S)^2$, and $(S_1,S_2,t)\in G\times(0,T]$ needs automatically to define an artificial boundary conditions on $\Gamma_0\subset\partial G=\{(S_1,S_2=\bar S),S_1\in(0,\bar S)\}\cup\{(S_1=\bar S,S_2),S_2\in(0,\bar S)\}$, which are given as usual

$$u(S_1,S_2,t)=0,\quad \forall(S_1,S_2)\in\Gamma_0,\ t\in(0,T]. \tag{20}$$

The next step, there is *discretization*, in particular, *time*, and *space discretization*.Let $0=t_0<t_1<\cdots<t_m=T$ be a set of discrete time points breaking $[0,T]$ into set of subintervals $\{[t_{k-1},t_k]\}$, $k=1,\ldots,m$. Hence, at any t_k, we get a function $u^k(S_1,S_2)=u(S_1,S_2,t_k)$. The time discretization of (15) means to approximate $\partial u(S_1,S_2,t)/\partial t$ by finite difference technique, e.g. by backward Euler scheme

$$\frac{1}{\Delta_k}\big(u^k(S_1,S_2)-u^{k-1}(S_1,S_2),v\big)_{L^2(G)}+a_{t_k}(u^k(S_1,S_2),v)=0,\ \forall v\in W_{\mathrm{b}},\ t_k,k=1,\ldots,m, \tag{21}$$

here $\Delta_k=t_k-t_{k-1}$, $k=1,\ldots,m$, and $u^0(S_1,S_2)=u_0(S_1,S_2)$, is the payoff function given, and $a_{t_k}(u^k,v)$ is given by bilinear form (16), calculated at $t=t_k$, precisely. *Space discretization* – Following general idea of FEM, the discretization with respect to $(S_1,S_2)\in G$ consists of covering G by a set of non-overlapping finite elements, e.g. triangles, where each of them carry a suitable function approximating the searched solution locally over an element. General motivating idea of FEM is to replace function space W_{b} by finite-dimensional sub-space $W_h\subset W_{\mathrm{b}}$, where $h>0$ accents a finite-dimensional nature of W_h conveniently, representing some quantitative characterictics of FE mesh, e.g. the area of the smallest element, the smallest vertex angle, etc. The simplest approximation function on a triangle is a linear function. The triangulation $\mathcal{T}_h$ of G provides a set of vertices $\mathcal{N}_h=\{N_i\}$, $i=1,\ldots,n$. At any vertex N_i one builts continuous piecewise linear approximation function $\psi_i(S_1,S_2)$, with a local support consisting with all triangles adjacent to N_i. The finite-dimensional sub-space W_h is

$$W_h=\mathrm{span}(\psi_1(S_1,S_2),\ldots,\psi_n(S_1,S_2)). \tag{22}$$

Emphasizing the triangulation $\mathcal{T}_h$, and thus the set of vertices $\mathcal{N}_h$, as well, do not depend upon time $t\in(0,T]$, we can express approximation of functions $u^k(S_1,S_2),k=1,\ldots,m$, in following way

$$u^k(S_1,S_2)\ \sim\ u^k_h(S_1,S_2)=\sum_{i=1}^n c_i^k\psi_i(S_1,S_2)\ \in W_h,\quad k=1,\ldots,m, \tag{23}$$

where $c_i^k,\ i=1,\ldots,n,\ k=1,\ldots,m$ are unknowns to be determined by solution of recurrent system of equations

$$\frac{1}{\Delta_k}\Big(\int_{\mathcal{T}_h}(u^k_h(S_1,S_2,t_k)-u^{k-1}_h(S_1,S_2,t_{k-1}))v_h\mathrm{d}S_1\mathrm{d}S_2\Big)+a_{t_k}(u^k_h,v_h)=0,\ k=1,\ldots,m,\ \forall v_h\in W_h. \tag{24}$$

with initial condition $u^0_h(S_1,S_2)=u_0(S_1,S_2)$, and (20), all being approximated within W_h.

Input data – We set $K=40,r=0.05,\sigma_1=0.3,\sigma_2=1.2,T=0.5$, and $S_1,S_2\in(0,65]$, i.e. $\bar S=65$, and 9 variants of $\rho=\{-1,-.5,0,.1,.3,.5,.7,.9,1\}$, with different σ_1,σ_2, intentionally.

The corresponding results for all variants are given in Figures 1, 2, and 3, thus showing noticeable changes of iso-band patterns of individual cases with corresponding ρ. The maximal iso-band values, i.e. $\max V(S_1,S_2,0)$, computed by our FreeFem++ code, are given in Table 2, and depicted in Figure 4, together with Eurepean put option prices for comparison computed for $\sigma=0.3$, and 1.2, too. The interesting is an increasing trend starting from $\rho=-1$, upto $\rho=1$, with a local peak at $\rho=0$. The open source sw FreeFem++ is fully described in [3], and [4].

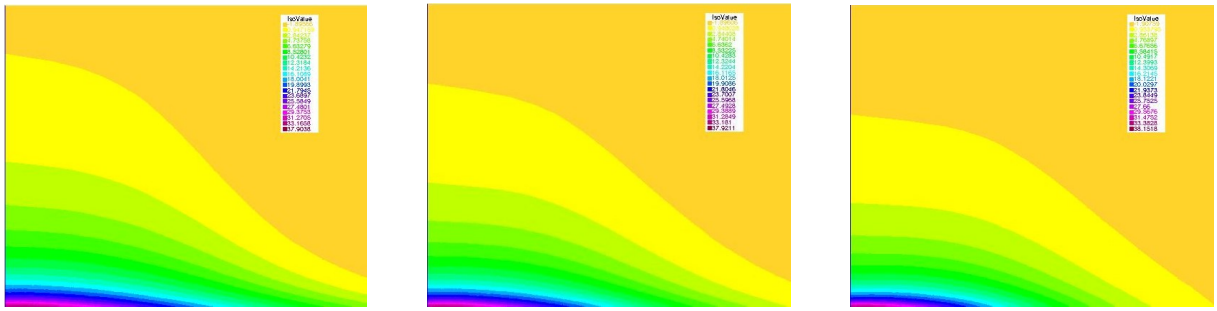


Figure 1: Meshes for cases: $\rho = -1.0\,(\text{left}),\quad -0.5\,(\text{middle}),\quad 0.0\,(\text{right})$

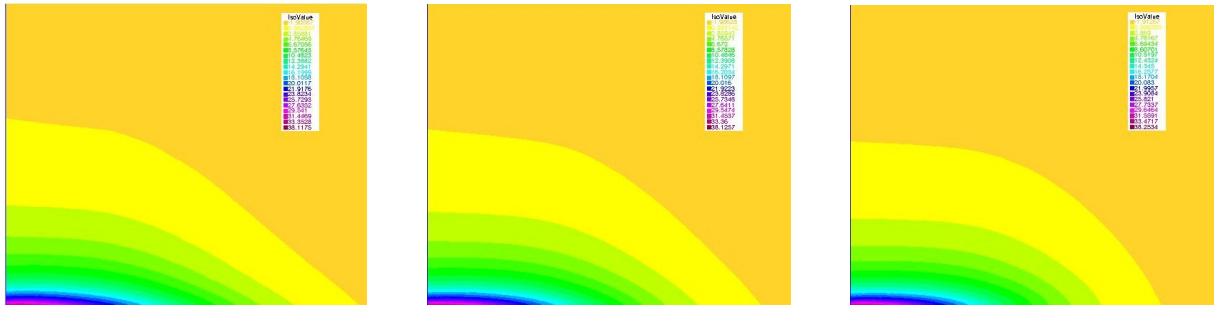


Figure 2: Meshes for cases: $\rho = 0.1\,(\text{left}),\quad 0.3\,(\text{middle}),\quad 0.5\,(\text{right})$

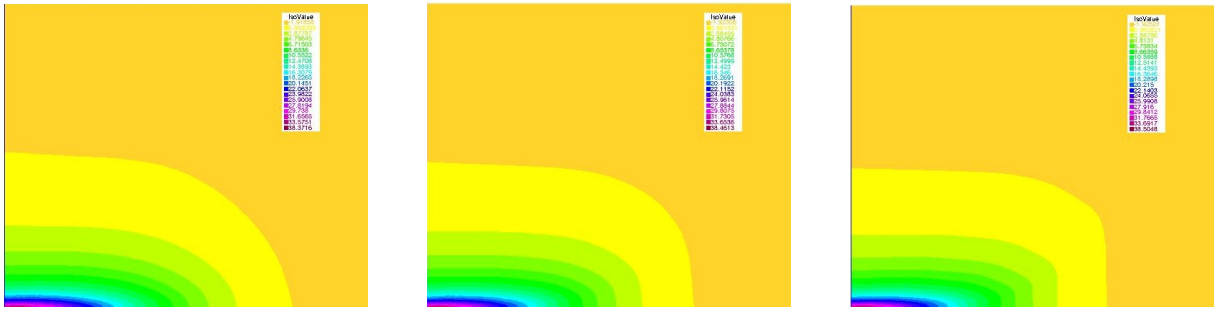


Figure 3: Meshes for cases: $\rho = 0.7\,(\text{left}),\quad 0.9\,(\text{middle}),\quad 1.0\,(\text{right})$

case:	1	2	3	4	5	6	7	8	9
ρ	-1.0	-0.5	0.0	0.1	0.3	0.5	0.7	0.9	1.0
$\max V(S_1, S_2, 0)$	37.904	37.921	38.152	38.118	38.128	38.253	38.372	38.461	38.505

Table 2: Case variants of ρ, and corresponding $\max V(S_1, S_2, 0)$

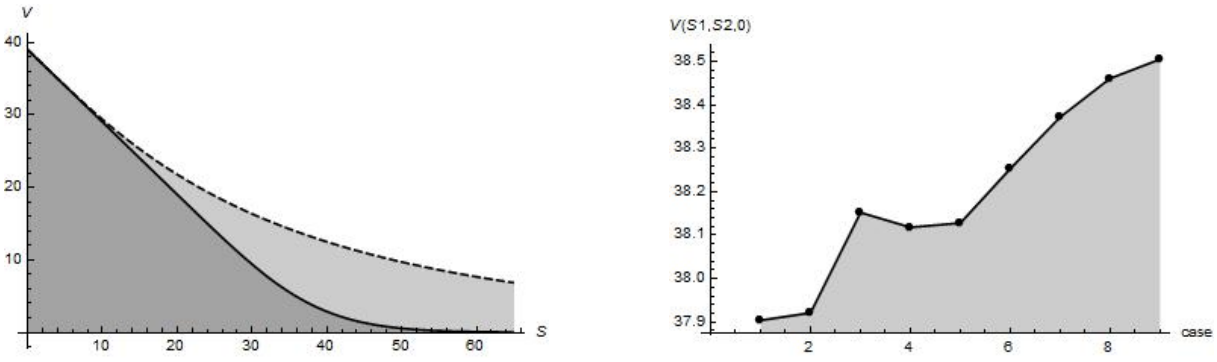


Figure 4: *left*: European put option prices $V_0(S), \sigma = .3$ (*full*), and $= 1.2$(*dashed*); *right*: maximal iso-band values $\max V(S_1, S_2, 0), i = 1, \ldots, 9$, given in Table 2.

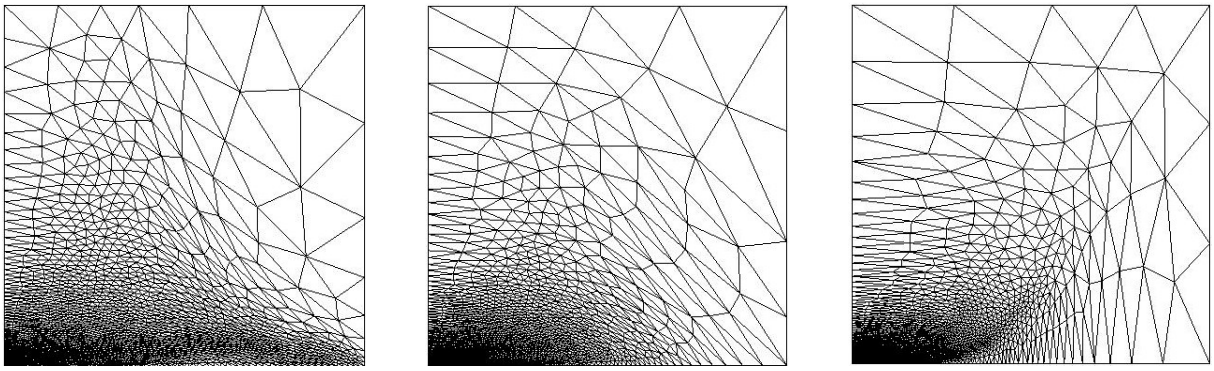


Figure 5: Triangulation meshes for cases: $\rho = -1.0\,(\text{left}),\ \ 0.0\,(\text{middle}),\ \ 1.0\,(\text{right})$

3 Conclusion

In the paper, we discussed briefly multi-asset European put option pricing problem formulation assuming spot prices of underlying asset to follow geometric Brownian motion with correlation structure. Delineating traditional approach based on self-financing portfolio and application of Itô's formula to option price, we obtained resulting partial differential equation describing the evolution of option price in space and time. Both classical and variational formulations are discussed. For demonstration purposes, we selected 2-D basket European put option pricing problem with the best-of put payoff function, and with different values of correlation coefficient ρ.

For discretization of the given problem, we use finite element method (FEM), and we sketched briefly both localization of the problem, i.e. truncation of an unbounded region into a bounded one, and its time and space discretization, as well.

For numerical solution of the problem, we use an open source sw package FreeFem++. Finally, we show all computed results which confirm the noticeable influence of correlation coefficient between underlying assets upon the 2-D basket European put option price. By the way, there is also rather interesting to recognize an influence of that correlation coefficient upon triangulation meshes, when intrinsic optimal meshing algorithm within the FreeFem++ package being invoked by the code.

Sure, there is a lot of computational experiments to perform, e.g. with different combinations among volatilities and correlation coefficient. Albeit the presented case study being limited in its extent, the promissing results support our intention and effort to continue our research in the field of computational finance.

References

[1] Achdou, Y., and Pironneau, O.: *Computational Methods for Option Pricing*. Frontiers in Applied Mathematics, vol. 30, SIAM, Philadelphia, 2005.

[2] Bossu, S.: *Advanced Equity Derivatives Volatility and Correlation*. John Wiley & Sons Inc., Hoboken, New Jersey, 2014.

[3] Hecht, F.: (2016). *FreeFem++*. 3-rd Ed., Version 3.51, *http://www.freefem.org*.

[4] Hecht, F.: New development in FreeFem++. *J. Numer. Math.* **20** (2012), no. 3-4, 251–265.

[5] Hilber, N., Reichmann, O., Schwab, Ch., and Winter, Ch.: *Computational Methods for Quantitative Finance – Finite Element Methods for Derivative Pricing.* Springer Finance, Springer, Heidelberg, 2013.

[6] Lukáš, L.: Variational Formulation of Option Pricing Problem a Platform for Finite Element Method in Finance. In: *Conf. Proceedings, 33-rd Int. Conf. Mathematical Methods in Economics 2015*, (Martinčík, D., Ircingová, J., and Janeček, P., eds.). Univ. of West Bohemia, Pilsen, 2015, 467–472.

[7] Shiryaev, A.N.: *Essentials of Stochastic Finance – Facts, Models, Theory*. World Scientific Publ. Co., Singapore, 2008.

Model of Passengers to Line Assignment in Public Municipal Transport

Tomáš Majer[1], Stanislav Palúch[2], Štefan Peško[3]

Abstract. The paper solves the situation of designing a completely new line system of a studied region. The goal is to build a new integrated transportation system containing several types of transportation lines like bus lines, tram lines, railway lines etc. A logical requirement is that the new system does not contain parallel bus, tram and railway lines.
In this paper we suppose that we are given an OD matrix describing transportation demand and moreover a line system with all possible lines. Several alternatives of shortest transport routes for every relation of OD matrix are calculated and passengers are assigned to these alternatives. The proposed model is based on method named PRIVOL proposed in [1] by Černý, Černá with the difference that demand on edges of transportation network is calculated by model itself.
The result of this calculation is determining capacities, number of trips resp. time intervals on proposed lines. The solution contains also identification of places in transportation network where facilities for changing lines should be built.

Keywords: public transport, shortest path, mixed linear programming

JEL classification: R41
AMS classification: 90B06

1 Introduction

Suppose we want to build a public personal transport system for a medium–sized region. We know the demand for transport specified by an origin–destination matrix $\mathbf{O} = (O_{ij})$. Individual elements O_{ij} reflect the number of passengers who need to travel over a given time period from the place i to the place j. The time period is morning or afternoon traffic peak, noon saddle or whole day during a Saturday, Sunday, or public holiday.

Let L be a bus line that is specified by the sequence of stops on which the vehicle stops $L = (i_1, i_2, i_3, \ldots)$. We have a set of lines $\mathcal{L} = \{L_1, L_2, L_3, \ldots\}$. The lines were operated in the region in the past or were designed by traffic engineers based on changes in residential areas, industrial zones, shop centers and transport infrastructure.

Our task is to design such a line system (subset of $\mathcal{L}$) to meet the demand and use a small number of vehicles. Of course, the public transport system must be attractive to the passengers and it is therefore necessary to minimize the passengers time lost.

Many mathematical models are known to try to ensure the goal mentioned above. We will improve model PRIVOL proposed by Černy, Černa in [1] and observed in detail in [2], [3]. The Černy's model divides the specified number of vehicles N so that demand D_a is satisfied on every arc a of the transport network. Variable x_l denotes the number of vehicles assigned to line L_l with the offer $F_l = C/T_l$, where C is capacity of vehicle and T_l is driving time including time needed to handle before and after driving. The criterion is to maximize the margin over which the offer exceeds the demand y. $\mathcal{L}_a$ denotes set of lines driving on arc $a \in A$.

$$\text{maximize} \quad y \tag{1}$$

$$\text{subject to} \quad \sum_{l \in \mathcal{L}} x_l = N \tag{2}$$

$$\sum_{l \in \mathcal{L}_a} \frac{F_l x_l}{D_a} \geq y, \ \forall a \in A \tag{3}$$

$$x_l \in \mathbb{N}_0 \quad \forall l \in \mathcal{L}, y \in \mathbb{R}_0^+. \tag{4}$$

[1]University of Žilina, Department of mathematical methods and operational research, Univerzitná 8215/1, 010 26 Žilina, Slovakia, tomas.majer@fri.uniza.sk
[2]University of Žilina, Department of mathematical methods and operational research, Univerzitná 8215/1, 010 26 Žilina, Slovakia, stanislav.paluch@fri.uniza.sk
[3]University of Žilina, Department of mathematical methods and operational research, Univerzitná 8215/1, 010 26 Žilina, Slovakia, stefan.pesko@fri.uniza.sk

The problem is how to calculate demand D_a on arc a of transport network. The easiest way is to find the shortest i–j path $\mu(i,j)$ for every non-zero element O_{ij} of OD matrix $\mathbf{O}$ and demand D_a calculate as sum of all passengers traveling over the arc a

$$D_a = \sum_{i,j|a\in\mu(i,j)} O_{ij}. \tag{5}$$

In our paper [4], we presented a method that counts K–shortest paths, and the passengers are proportionally distributed between them. However, this method did not provide satisfactory results because passengers do not always have to use the shortest path in the transport network, but they will adapt to the existing lines and the timetable. In this article we present the way in which we assign passengers not only to the individual arcs of the transport network but also to individual lines operating at these arcs. Model (1)–(4) is used to make this assignment optimal.

2 Proposed model

2.1 Extended model of transport network

We first create a public transport model consisting of several layers. The lowest layer will represent the arrival of passengers at the stop and their boarding on the public transport vehicle. The top layer will represent the exit of passengers from the public transport vehicle and the termination of their journeys. The middle layers will represent individual lines.

So we create an arc weighted directed graph $G = (V, A)$ with set of vertices $V = \{(l,i)|l \in \mathcal{L} \cup \{\text{boarding}, \text{exit}\}, i \in S\}$, where S is set of stops. Set of arcs A contains arcs of

- boarding to line l on stop i: $((\text{boarding}, i), (l,i))$,
- exit from line l on stop j: $((l,j), (\text{exit}, j))$,
- traveling by line l from stop i to stop j: $((l,i), (l,j))$,
- move from line l_1 to line l_2 on stop i: $((l_1,i), (l_2,i))$.

The weight of arc is traveling time or time need for boarding, exit or move between lines. Every $(\text{boarding}, i)$–(exit, j) path represent traveling from stop i to stop j using one or more lines and the length of the path is total time needed to move. Several paths represents alternatives of traveling.

Let's assume that for each pair of stops i, j for which $O_{ij} > 0$, we have a few of such paths precalculated. We denote these paths $\mu_k(i,j)$.

2.2 Mathematical model of problem

Let the number of passengers traveling from stop i to stop j using path $\mu_k(i,j)$ is equal to z_{ijk}. Then

$$\sum_{k|\exists\mu_k(i,j)} z_{ijk} = O_{ij} \qquad \forall i,j \in S. \tag{6}$$

Transport demand on line l through arc a can be calculated as sum of number of passengers using any path containing arc a.

$$d_{l,a} = \sum_{\substack{i,j,k \\ a\in\mu_k(i,j)}} z_{ijk} \qquad \forall l \in \mathcal{L}, \forall a = ((l,u)(l,v)) \in A. \tag{7}$$

Finnaly, we can combine model (1)–(4) and constraints (6), (7). Because we need to preserve linearity of model, we fix value y to some constant Y and change criteria to minimize number of used vehicles.

$$\text{minimize} \qquad n = \sum_{l \in \mathcal{L}} x_l \tag{8}$$

$$\text{subject to} \qquad F_l x_l \geq Y d_{l,a}, \qquad \forall l \in \mathcal{L},\ \forall a = ((l,u)(l,v)) \in A \tag{9}$$

$$\sum_{\substack{i,j,k \\ a \in \mu_k(i,j)}} z_{ijk} = d_{l,a} \qquad \forall l \in \mathcal{L}, \forall a = ((l,u)(l,v)) \in A \tag{10}$$

$$\sum_{k | \exists \mu_k(i,j)} z_{ijk} = O_{ij} \qquad \forall i,j \in S \tag{11}$$

$$x_l \in \mathbb{N}_0,\ z_{i,j,k} \in \mathbb{N}_0,\ d_{l,a} \in \mathbb{R}_0^+. \tag{12}$$

3 Illustrative example

We have a transport system with 8 stops and 3 lines as shown in Figure 1.

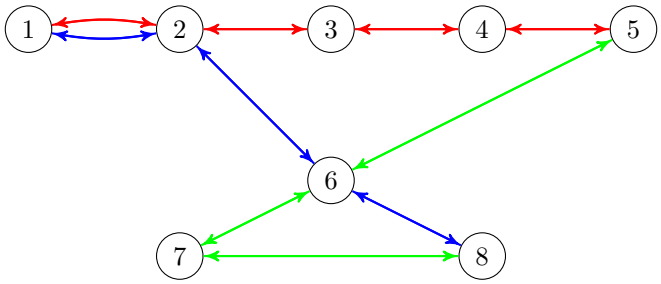


Figure 1 Diagram of transport network and bus lines

The set of lines $\mathcal{L} = \{A, B, C\}$. Because each line has two directions - forward and backward. Line A_f is sequence of stops $A_f = (1, 2, 3, 4, 5)$, line A_b is sequence of stops $A_b = (5, 4, 3, 2, 1)$. Line B_f is sequence of stops $B_f = (5, 6, 7, 8)$, line B_b is sequence of stops $B_b = (8, 7, 6, 5)$. Line C_f is sequence of stops $C_f = (1, 2, 6, 8)$, line C_b is sequence of stops $C_b = (8, 6, 2, 1)$.

Extended transport network $G = (V, A)$ consists of vertices

$$\begin{aligned} V = \{&(\text{board}, 1), (\text{board}, 2), (\text{board}, 3), (\text{board}, 4), (\text{board}, 5), (\text{board}, 6), (\text{board}, 7), (\text{board}, 8), \\ &(A,1), (A,2), (A,3), (A,4), (A,5), (B,5), (B,6), (B,7), (B,8), (C,1), (C,2), (C,6), (C,8), \\ &(\text{exit}, 1), (\text{exit}, 2), (\text{exit}, 3), (\text{exit}, 4), (\text{exit}, 5), (\text{exit}, 6), (\text{exit}, 7), (\text{exit}, 8)\} \end{aligned}$$

and arcs

$$\begin{aligned} A = \{&((\text{board}, 1), (A, 1)), (\text{board}, 1), (C, 1)), ((\text{board}, 2), (A, 2)), ((\text{board}, 3), (A, 3)), ((\text{board}, 4), (A, 4)), \\ &((\text{board}, 5), (A, 5)), (\text{board}, 5), (B, 5)), ((\text{board}, 6), (B, 6)), ((\text{board}, 6), (C, 6)), ((\text{board}, 7), (B, 7)), \\ &((\text{board}, 8), (B, 8)), (\text{board}, 8), (C, 8)), \\ &((A,1),(A,2)), ((A,2),(A,3)), ((A,3),(A,4)), ((A,4),(A,5)), \\ &((A,5),(A,4)), ((A,4),(A,3)), ((A,3),(A,2)), ((A,2),(A,1)), \\ &((B,5),(B,6)), ((B,6),(B,7)), ((B,7),(B,8)), ((B,8),(B,7)), ((B,7),(B,6)), ((B,6),(B,5)), \\ &((C,1),(C,2)), ((C,2),(C,6)), ((C,6),(C,8)), ((C,8),(C,6)), ((C,6),(C,2)), ((C,2),(C,1)), \\ &((A,1),(\text{exit},1)), ((C,1),(\text{exit},1)), ((A,2),(\text{exit},2)), ((A,3),(\text{exit},3)), ((A,4),(\text{exit},4)), \\ &((A,5),(\text{exit},5)), ((B,5),(\text{exit},5)), ((B,6),(\text{exit},6)), ((C,6),(\text{exit},6)), ((B,7),(\text{exit},7)), \\ &((B,8),(\text{exit},8)), ((C,8),(\text{exit},8)), \\ &((A,1),(C,1)), ((C,1),(A,1)), ((A,2),(C,2)), ((C,2),(A,2)), ((A,5),(B,5)), ((B,5),(A,5)), \\ &((B,6),(C,6)), ((C,6),(B,6)), ((B,8),(C,8)), ((C,8),(B,8))\} \end{aligned}$$

It should be noted that the segment between stops 1 and 2 is served by different lines A and C and therefore there are different vertices $(A, 1)$, $(C, 1)$, $(A, 2)$ and $(C, 2)$ in our model. The section from stop 1 to stop 2 is represented by arc $((A, 1), (A, 2))$ (traveling by line A) and arc $((C, 1), (C, 2))$ (traveling by line C). Number of

passengers $d_{l,a}$ is calculated for all the lines and arcs separately in equation (7). The margin over which the offer exceeds the demand is validated for all the lines and arcs separately by condition (9) of our model too. The total number of passengers traveling between two stops by any line is not required in our model.

We can omit the time needed for boarding and exit, as there is just one boarding and just one exit in each way. Suppose that time to move between lines is allways 5 minutes and traveling time between two subsequent stops on every line is 2 minutes. So we can find three 3–8 paths:

$$\mu_1(3,8) = ((\text{board},3),(A,3),(A,4),(A,5),(B,5),(B,6),(B,7),(B,8),(\text{exit},8)),\ \text{len}(\mu_1(3,8)) = 15,$$
$$\mu_2(3,8) = ((\text{board},3),(A,3),(A,4),(A,5),(B,5),(B,6),(C,6),(C,8),(\text{exit},8)),\ \text{len}(\mu_1(3,8)) = 18,$$
$$\mu_3(3,8) = ((\text{board},3),(A,3),(A,2),(C,2),(C,6),(C,8),(\text{exit},8)),\ \text{len}(\mu_1(3,8)) = 11,$$

It follows that

$$z_{3,8,1} + z_{3,8,2} + z_{3,8,3} = O_{3,8}$$

The arc $a_{15} = ((A,3),(A,4))$ is in $\mu_1(3,8)$ and $\mu_2(3,8)$, but not on path $\mu_3(3,8)$. So

$$d_{A,a_{15}} = z_{3,8,1} + z_{3,8,2} + \sum_{\substack{i,j,k \\ i\neq 3, j\neq 8 \\ a_{15}\in\mu_k(i,j)}} z_{ijk}.$$

If we set handle time to 5 minutes, then $T_A = 2\cdot(4\cdot 2+5) = 26$ minutes. Let capacity of vehicle be $C = 80$, then one vehicle can take $80\cdot 60/26$ passengers on every arc in one hour and therefore

$$80\cdot 60/26\cdot x_A \geq Y\cdot d_{A,a_{15}}.$$

To build whole model with all conditions, K–shortest paths between each pair of stops must be calculated. This is an exhaustive work for man and so need to by programmed and calculated using computer.

4 Conclusion

In the paper, we suggested how to assign passengers in public municipal transport to lines with respect to demand described by origin–destination matrix and set of possible lines given. Passengers may move between lines during traveling and lines are served by minimal number of vehicles. From number of vehicles is easy to compute number of trips or time interval of subsequent trips on lines for regular interval timetable.

In our next work we would like to try our model on real data from selected self–governing territorial unit of Slovakia.

Acknowledgements

The authors are pleased to acknowledge the financial support of the Slovak research and development agency under the grant No. APVV-14-0658 “Optimization of urban and regional public personal transport”.

References

[1] Černá, A., Černý, J.: Linear Optimisation of Urban Bus Routes and Frequencies. *Czechoslovak Journal for Operational Research* **vol. 1** (1992), 207-217

[2] Černý, J., Černá, A.: Erlander Principle in Managerial Decision Making on Czech and Slovak Urban Transport Routes. *E+M Ekonomie a Management* **vol. 16**, 1 (2013), 93–100.

[3] Daněk J., Plevný M., Teichmann D.: The Rational Operation of the Urban Transport Line Network by Minimization of the Needed Vehicles. *E+M Ekonomie a Management* **vol. 13**, 2 (2010), 53–61.

[4] Majer T.: Route Assignment Based on k-Shortest Paths Problem *QME 2010, 15th international conference, Smolenice, Slovakia* (2010), 117–126.

Evaluating the riskiness of hedging strategy with gold, US dollar and Czech crown

Jiří Málek[1], Quang Van Tran [2]

Abstract. This article evaluates the impact of securing dollar deposits using gold from the view of a foreign currency (Euro, CZK) holder. Alpha stable distribution is used as an approximation for log-return distribution and its parameters are estimated by maximum likelihood estimation method. Despite the negative correlation between the gold price and the exchange rate, diversification negligibly affects the value of alpha (fat tails) and beta (skewness) coefficients. However, the diversification effect works as the calculated VaR99 and CVaR97.5 values of the corresponding portfolios show. Measuring by this metric, the risk is also the smallest for portfolios with their weight of gold at 40%.

Keywords: alpha stable distribution, USD/EUR, USD/CZK, gold, VaR, CVaR

JEL classification: 91B82, 91B28
AMS classification: G10, G120

1 Introduction

It is known that the gold price (in USD) is negatively correlated with the dollar exchange rate against the major world currencies. That is why gold is often used as a security against the decline of the dollar exchange rate. If an investor (in the Eurozone or in the Czech Republic) owns a dollar deposit, he tries to secure himself against the exchange rate by purchasing gold. Since the negative correlation is changing and sometimes even positive, we may want to know about the impact of this security on the profitability of this portfolio (gold and dollars in terms of foreign currency). In this paper, to answer this question the alpha stable distribution, whose usefulness in finance can be found in [6], is used as an approximation of log-return distribution and its parameters is examined for portfolios with different gold and currency ratios. In the first part, the basic characteristics of the stable distribution are presented. In the second part, the gold price behavior is briefly described. After that, in the following empirical part, we perform parameter estimation of the distribution using MLE for different portfolios. Then VaR, CVaR 97.5[3] and other riskiness indicators of returns alpha stably distributed are also calculated and compared.

2 Alpha stable distribution

Alpha stable distribution is characterized by its characteristic function $\phi_0(t)$

$$\log \phi_0(t) = \begin{cases} -\sigma^\alpha \mid t \mid^\alpha \{1 + i\beta \text{sign}(t) \tan \frac{\alpha\pi}{2}[(\sigma \mid t \mid)^{(1-\alpha)} - 1]\} + i\mu_0 t, \ \alpha \neq 1 \\ -\sigma \mid t \mid \{1 + i\beta \text{sign}(t) \frac{2}{\pi} \log(\sigma \mid t \mid)\} + i\mu_0 t, \ \alpha = 1 \end{cases} \tag{1}$$

There are four parameters $\alpha, \beta, \mu, \sigma$ included in this characteristic function. α, β are shape parameters, where $\alpha \in (0, 2]$ is the tail power as α decreases tail thickness increases and $\beta \in [-1, 1]$ is skewness parameter. The meaning of two remaining parameters is: μ is location parameter (generalized mean) and σ is scale parameter (generalized variance). A variance does not exist for $\alpha \in (0, 2)$ and a mean for $\alpha \in (0, 1]$ (for more detailed information on stable distribution, see [1, 2, 3]).

The following properties holds for a stable distribution:
1) Let X_1, X_2 are the independent stable random variables, with $X_i = S(\alpha, \beta_i, \sigma_i, \mu_i)$, $i = 1, 2$, then

$$X_1 + X_2 \sim S(\alpha, \beta, \sigma, \mu) \ \text{ with} \tag{2}$$

[1]University of Economics in Prague, Department of Banking and Insurance, nám. W. Churchilla 4, Praha 3, Czech Republic, GMalek@post.cz

[2]University of Economics in Prague, Department of Statistics and Probability, nám. W. Churchilla 4, Praha 3, Czech Republic, tran@vse.cz

[3]Bank for International Settlements recommended CVaR97.5 because VaR99 is equal to the CVaR97.5 for a normal distribution.

- $$\sigma = \sigma_1 + \sigma_2 \tag{3}$$
- $$\beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha} \tag{4}$$
- $$\mu = \mu_1 + \mu_2 \tag{5}$$

2) If $X \sim S(\alpha, \beta, \sigma, \mu)$ and $a \in R$ then

$$X + a \sim S(\alpha, \beta, \sigma, \mu) + a \tag{6}$$

3) If $X \sim S(\alpha, \beta, \sigma, \mu)$ and $a \in R$ then

$$aX \sim S(\alpha, \text{sgn}(a)\beta, |a|\sigma, a\mu) \tag{7}$$

If X is a standardized stable variable ($\sigma = 1, \mu = 0$) with $\alpha \in (0, 2)$, then as $x \to \infty$

$$P(X > x) \approx (1 + \beta) C_\alpha x^{-\alpha} \tag{8}$$

where

$$C_\alpha = \left(2 \int_0^\infty x^{-\alpha} \sin x \right)^{-1} = \frac{1}{\pi} \Gamma(\alpha) \sin\left(\frac{\alpha \pi}{2} \right) \tag{9}$$

and

$$P(X < -x) \approx (1 - \beta) C_\alpha x^{-\alpha} \tag{10}$$

Here $f(x) \approx g(x)$ means $\lim_{x \to \infty} f(x)/g(x) = 1$.
That means that when $\beta > 0$ the right tail is heavier than the left one (but with same order α) and conversely. We remind that for standard normal distribution the tail probability as $x \to \infty$ is

$$P(X > x) \approx \frac{\exp(-x^2/2)}{x\sqrt{2\pi}} \tag{11}$$

3 Gold versus dollar

Figure 1: Trade weighted USD index versus gold price (Source: Federal Reserve LBMA)

There is an inverse relationship between the trade-weighted U.S. dollar[4] and the gold price (Fig. 1). Trade-weighted value shows how the U.S. dollar is gaining or losing purchasing power—compared to its trading partners.

[4]The trade-weighted US dollar index is a measure of the value of the United States dollar relative to other world currencies. Currently, this index is calculated by factoring in the exchange rates of six major world currencies: the euro, Japanese yen, Canadian dollar, British pound, Swedish krona and Swiss franc. The euro holds the most weight versus the dollar in the index, constituting about 58 percent followed by the yen with about 14 percent.

However, this inverse relationship is not as precise as it used to be under the gold standard. Even though the gold standard is gone, there is still a psychological tilt towards gold whenever the value of the U.S. dollar decreases. The inverse relationship remains because a falling dollar increases the value of other countries' currencies. This increases the demand for commodities including gold. It also increases the prices. And when the U.S. dollar starts to lose its value, investors look for alternative investment sources to store value. Gold is an alternative.

However it is possible for the U.S. dollar and gold price to increase at the same time. This can occur because of a crisis in some other country or region. Then investors buy safer assets as the U.S. dollar and gold. The U.S. dollar is also driven by many factors—like monetary policy and inflation in the U.S. vs. other countries. It is also driven by economic prospects in the U.S. vs. other countries. Investors need to consider all of these factors.

The inter-market analysis often suggests to do investments based upon various correlations between different markets. That means that when one market moves in one direction, we should see a correspondingly opposite move in an inversely "correlated" market. Many have used the bond market in such a manner in relation to the equity market, and many have used the (USD) relative to the gold market. But the problem with using such seeming "correlations" is that they can disappear just as quickly as they appear. Over the last few years, we have seen such decoupling in many markets. Gold and the U.S. dollar were associated when the gold standard was being used. During this time, the value of a unit of currency was tied to the specific amount of gold. The gold standard was used from 1900 to 1971. The separation was made in 1971. They began to be valued based on supply and demand. The U.S. dollar became a fiat currency that gets its value from government regulation, but is not backed by a physical commodity. It is traded on foreign markets and was used as a reserve currency. Gold moved to floating exchange rates after 1971. This made its price vulnerable to the U.S. dollar's external value. In 2008, the International Monetary Fund (or IMF) estimated that 40–50 percents of the moves in the gold prices since 2002 were dollar-related.

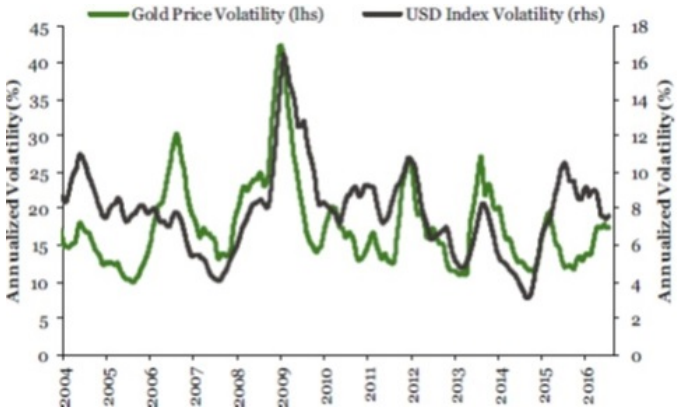


Figure 2: Volatility of gold price and dollar index (Source: Bloomberg)

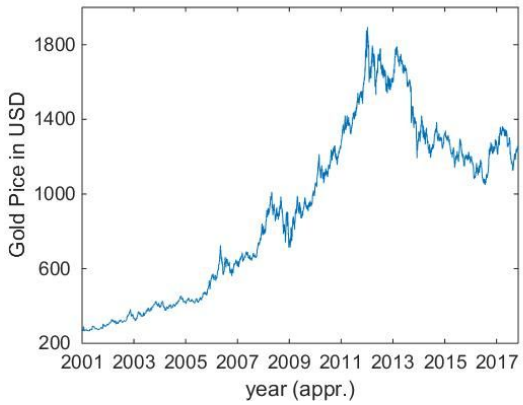


Figure 3: The development of gold price

The competitive devaluations resulting from central bank policy activities has driven currency volatility higher, which has in turn flowed into other asset markets. USD volatility has a strong historical relationship with gold volatility (Fig. 2). The strength of the correlation of gold and USD volatility highlights the 'currency' characteristics of gold. The divergence during 2015 highlights the varied role that gold plays within investment portfolios. At a time of rising USD volatility as the market was concerned about rate hikes and Chinese market contagion, gold volatility moderated and prices softened over the ensuing six months.

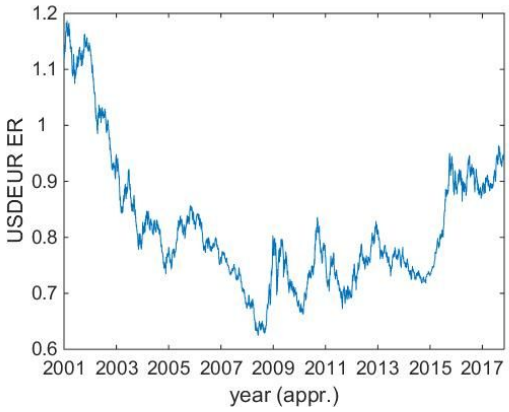


Figure 4: The development of USD/EUR exchange rate

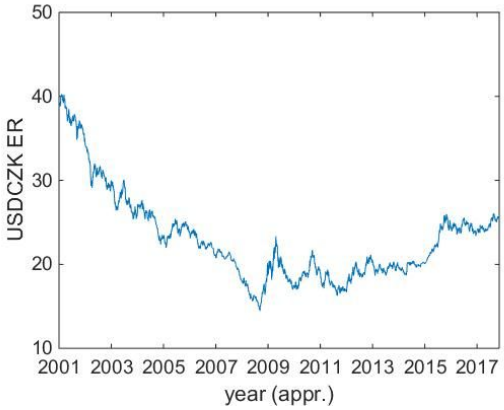


Figure 5: The evolution of USD/EUR exchange rate

4 Empirical analysis and results

For our analysis, three series gold price, USD/EUR and USD/CZK exchange rate series are used. They are daily data from 5.2001 to 2.2017. The development of gold price and USD/EUR exchange rate in this period are displayed in Fig. 3 and 4 of USD/CZK exchange rate on Fig. 5. The original data series are converted into the so called logarithmic returns series. The descriptive statistics of original series as well as their corresponding returns are shown in Table 1. We also calculate correlation coefficient of gold price and USD/EUR exchange rate and gold price returns and USD/CZK exchange rate returns for a rolling window with fixed length of 125 which corresponds to a half of a year trading period. The result is shown in Fig. 6. It is clear that their correlation is very unsteady over this period as it confirms what has been stated before.

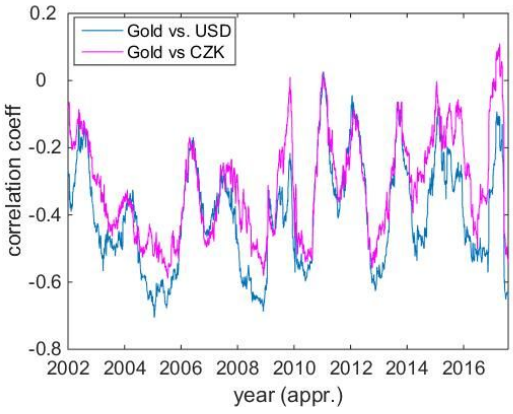


Figure 6: The evolution of correlation coefficient

Characteristic	Original series			Returns series		
	USD/CZK	USD/EUR	GOLD	USD/CZK	USD/EUR	GOLD
Mean	23.04	0.82	922.12	-0.0001	-0.00004	0.0004
Median	21.92	0.78	935.00	-0.0002	-0.0001	0.0006
Maximum	40.29	1.19	1895.00	0.053	0.030	0.0006
Minimum	14.45	0.62	263.95	-0.057	-0.046	-0.096
Std. Dev.	5.26	0.12	456.32	0.0079	0.0064	0.0119
Skewness	1.24	1.20	0.10	0.23	-0.06	-0.34
Kurtosis	4.35	4.10	1.73	7.30	5.37	7.58
Observations	3787	3787	3787	3786	3786	3786

Table 1: Descriptive statistics of analyzed data

First, we use the returns series to estimate the four parameters of α-stable distribution using maximum likelihood estimation method, see [5]. The estimation procedures are performed in Matlab. The estimation results of α and β for different portfolios are reported in Tables 2 and 3. Using α-stable distribution for returns of these portfolios, we calculate VaR for confidence levels 97.5% and 99% and CVaR for confidence level 97.5% (the detailed description of how they are calculated can be found in [4]). The results are reported in Table 4. Further, we compare risk measures and Sharpe ratios using the standard deviation and scale parameter estimated for normal distribution and α-stable distribution respectively. The results are displayed in Tables 5 and 6. The obtained results clearly show that the risk of a portfolio can be reduced by appropriate diversification using assets with negatively correlated returns.

Gold weight	α	SE α	p-value	β	SE β	p-value
0.000	1.800	0.024	0.000	0.160	0.013	0.000
0.200	1.793	0.025	0.000	0.253	0.096	0.008
0.400	1.734	0.016	0.000	0.091	0.008	0.000
0.600	1.646	0.032	0.000	-0.045	0.044	0.305
0.800	1.648	0.026	0.000	-0.093	0.018	0.000
1.000	1.669	0.050	0.000	-0.129	0.047	0.006

Table 2: Estimation results for alpha and beta, portfolio Gold-Euro

Gold weight	α	SE α	p-value	β	SE β	p-value
0.000	1.729	0.086	0.000	0.134	0.017	0.000
0.200	1.727	0.018	0.000	0.116	0.007	0.000
0.400	1.731	0.061	0.000	0.006	0.020	0.770
0.600	1.659	0.031	0.000	-0.073	0.042	0.083
0.800	1.654	0.028	0.000	-0.105	0.049	0.031
1.000	1.669	0.050	0.000	-0.129	0.047	0.006

Table 3: Estimation results for alpha and beta, portfolio Gold-CZK

Measure	Portfolio	0	0.2	0.4	0.6	0.8	1.0
VaR99	G+Euro	-0.0162	-0.116	-0.0129	-0.0207	-0.0292	-0.0380
VaR99	G+CZK	-0.0212	-0.0160	-0.0153	-0.0216	-0.0295	-0.0380
VaR97.5	G+Euro	-0.0123	-0.0086	-0.0090	-0.0132	-0.0185	-0.0243
VaR97.5	G+CZK	-0.0150	-0.0113	-0.0106	-0.0138	-0.0187	-0.0243
CVaR97.5	G+Euro	-0.0200	-0.0145	-0.0168	-0.0280	-0.0390	-0.0497
CVaR97.5	G+CZK	-0.0273	-0.0208	-0.0201	-0.0290	-0.0393	-0.0497

Table 4: Calculated VaR and CVaR

Gold weight		0	0.2	0.4	0.6	0.8	1
St. deviation	N	0.0064	0.0048	0.0048	0.0066	0.0091	0.0119
Sigma	Stable	0.0040	0.0029	0.0028	0.36	0.0050	0.0066
Mean$(*10^{-3})$		-0.0400	0.0480	0.1360	0.2240	0.321	0.4000
Sharpe$(*10^{-3})$	N	-0.0063	0.0100	0.0283	0.0339	0.0343	0.0336
Sharbe$(*10^{-3})$	Stable	-0.0100	0.0166	0.0486	0.0622	0.0624	0.0606

Table 5: Calculated volatility, mean and Sharpe ratio for gold and Euro portfolios

Gold weight		0	0.2	0.4	0.6	0.8	1
St. deviation	N	0.0079	0.0060	0.0056	0.0069	0.0091	0.0119
Sigma	Stable	0.0046	0.0034	0.0032	0.0038	0.0050	0.0066
Mean$(*10^{-3})$		-0.1000	0.0000	0.1000	0.2000	0.3000	0.4000
Sharpe$(*10^{-3})$	N	-0.0127	0.0000	0.0179	0.0290	0.0330	0.0336
Sharpe$(*10^{-3})$	Stable	-0.0217	0.0000	0.0313	0.0526	0.0600	0.0606

Table 6: Calculated volatility, mean and Sharpe ratio for gold and CZK portfolios

5 Conclusion

It turns out that, despite the negative correlation between the gold price and the exchange rate, diversification negligibly affects the value of the alpha and beta coefficients of stable distribution. For USD / Euro and USD / CZK, alpha is the highest for a pure investment in dollar deposits (0% gold). The maximum beta (positive skewness that most investors prefer) occurs for USD / Euro in the portfolio included gold with a weight of 20%, for USD / CZK is the beta highest for a pure investment in dollar deposits. In addition, if we include 60% of gold in USD / EUR deposit the left tail is the thickest (alpha = 1.648), for USD / CZK deposits the left tail is the thickest at 80On the other, the diversification effect works as it is shown by calculated values of VaR and CVaR. For USD / EUR, both values for VaR99 and CVaR97.5 are highest at a 20% gold and at USD / CZK at 40%. The risk is the smallest (measured by standard deviation resp. Sigma) is the smallest at a weight of 40% for both deposits. The Sharpe Index is the highest for a pure gold investment, but this is consequence the average return was positive for gold and negative for both currencies. As a conclusion, the recommended investment for USD/EUR the weight of gold in the portfolio may be around 20% (the best values of VaR 99 and CVaR, positive beta, low sigma risk). For CZK, the gold weight should be approximately 40% (the best VaR and CVaR, positive beta, sigma the lowest).

Acknowledgements

The support from the Czech Science Foundation under Grant P402/12/G097 is gratefully acknowledged.

References

[1] Belov, A.: On the Computation of the Probability Density Function of Stable Distribution Chambers, *Mathematical Modelling and Analysis, Proceedings of the 10th International Conference MMA, Trakai*, 2005. 333–341.

[2] Borak, S, Hardle, W. and Weron, R: Stable Distributions. *SFB 649 Discussion Paper 2005-008*, Center for Applied Statistics and Economics, Humbolt-Universitat zu Berlin, 2005

[3] Borak, S., Misiorek, A. and Weron, R.: Models for Heavy-tailed Asset Returns, *MPRA paper No. 25494*, September 2010.

[4] Dowd, K.: *Measuring Market Risk.* John Wiley & Sons Inc., West Sussex, 2002.

[5] Málek, J. and Tran van Quang: Evaluating the ability to model the heavy tails of asset returns of two alternative distributions, *Mathematical Methods in Economics, Proceedings of the 34th International Conference, Liberec, Czech Republic*, September 6th – 9th, pp. 528-533, 2016.

[6] Nolan, J. P.: Modeling Financial Data with Stable Distribution. In: *Handbook of Heavy Tailed Distributions in Finance*, (S. Rachev, ed.), Elsevier Science Amsterdam 2003, 105–130

Achilles and the Tortoise – the Story of the Minimum Wage

Lubos Marek[1], Michal Vrabec[2]

Abstract. The question of the minimum wage has often been discussed recently. On the one hand, there are some politicians who want to raise it, which is not surprising with the elections approaching. On the other hand, economic experts, and especially employers, are reluctant. Let us, for the moment, accept that the minimum wage is to be raised; a question remains by how much, or on which economic variables its increase should be dependent. One possibility is to tie the minimum wage with the average wage. The present paper will try to show that this would be an absurd option. We are going to show on real data what the consequences of the increased minimum wage and, especially, the impact on the wage distribution would be. Further, we will consider different levels of the minimum wage and study several scenarios related to changes in all wages after an increase of the minimum wage limit.

Keywords: wage, minimum wage, wage characteristics, wage distribution.

JEL Classification: C44
AMS Classification: 90C15

1 Minimum Wage

The topic of the minimum wage has been widely discussed in the media recently and has acquired political connotations due to the approaching elections. In the present paper, we will steer clear of the political side of the minimum wage; neither will we consider the full economic impacts of its increase. We will consider and analyze a few scenarios to illustrate the consequences for the wage distribution. Further we will show that the effort to tie the increase of the minimum wage with the value of the average wage is a never-ending story, similar to the classical example of Achilles and the tortoise. We will study the empirical distribution of wages and its response to different scenarios of a minimum wage increase. This empirical distribution can be described with the aid of a theoretical probabilistic model – cf. [3] or [5]. We will mainly be interested in the basic statistical characteristics – the location (average and quantiles), the variability expressed by the variance, and the skewness and kurtosis – cf., e.g., [4].

Before analyzing the data, we will have a look at the idea of tying the increase of the minimum wage with the value of the average wage – more or less regardless of which actual formula would be used for that increase. Let us recall the well-known paradox of Achilles and the tortoise.

The Eleatic Zeno is said to have claimed a proof for impossibility of movement. Achilles – the fastest runner – will never overtake the tortoise in front of him. When he reaches the original position of the tortoise, it has already moved a small distance from that position. Achilles runs over that small distance, the tortoise has moved a bit further, and so on to infinity. Achilles' movement can thus be described as an infinite series of ever shortening segments, and this paradox was insurmountable for ancient Greek philosophers.

Aristotle in Metaphysics [1] shows the error of the paradox: a sum of an infinite series may be finite if the terms are decreasing quickly enough; which is the case here. Zeno's paradox was one of the sources from which infinitesimal calculus eventually arose. An illustration of the problem is shown in Figure 1.

How far is this classical example from our problem concerning the minimum wage? As soon as the minimum wage is increased to x% of the average wage, the latter will necessarily be increased as well. If it is so, the minimum wage should be increased correspondingly, etc., etc. Hence a better methodology should be established for increasing the minimum wage, or stipulate that it will only be increased once a year (or once in two years, or three – what is economically feasible, after all?).

[1] University of Economics, Prague, department of Statistics and Probability, W. Churchill Sq.4, 130, 57 Prague 3, marek@vse.cz.
[2] University of Economics, Prague, department of Statistics and Probability, W. Churchill Sq.4, 130, 57 Prague 3, vrabec@vse.cz.

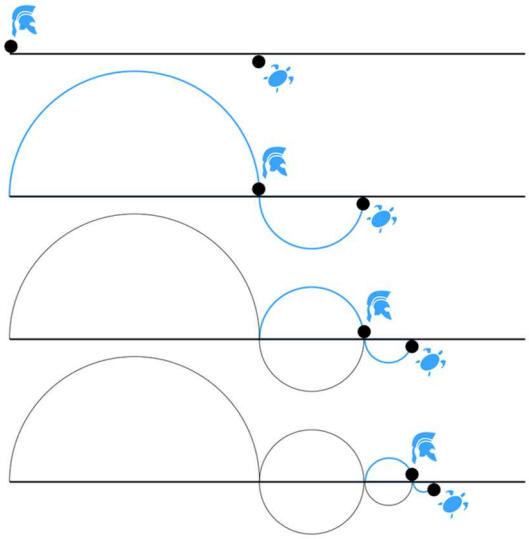


Figure 1 Achilles and the tortoise

2 Wage Distribution in the Czech Republic

First of all, let us have a look at the empirical distribution of wages in the Czech Republic in 2016.

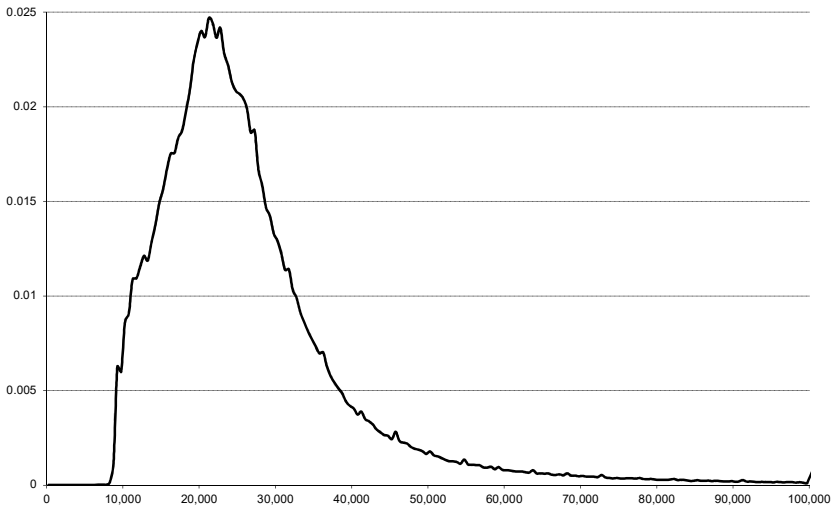


Figure 2: The empirical distribution of wages in the Czech Republic in 2016

It shows the polygon of wages in the Czech Republic in 2016. The curve is cut off at a value of 100,000 CZK. Approximately 1% of wages are above that level, which is insignificant for our modeling. We will mainly be interested in the left-hand side of the picture, in which low wages are situated, and of course the minimum wage as well – this fact is clearly seen from the chart. The curve is not smooth because it depicts the empirical data. Smoothing may be achieved by applying a probabilistic model. The wage distribution may be modeled with the aid of different probabilistic distributions. Log-normal distribution has been used for this purpose for many years. Modeling of wages has been addressed by many authors – e.g., [2] or [3]. The authors of the present paper have shown (Marek a Vrabec in [4]) that a suitably chosen mixture of log-normal distributions is a far better model. However, if we want to avoid using mixtures, the best solution may be the three-parametric logistic distribution with the probability density

$$f(x)=\frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}\left(1+\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{-2}, \qquad \gamma \le x < \infty \tag{1}$$

where the parameters have the following meanings: α – continuous shape parameter, β – continuous scale parameter, γ – continuous location parameter ($\gamma = 0$ yields the two-parameter log-logistic distribution). Estimates of these parameters of the theoretical distribution (Figure 3) are: $\alpha = 4.164$, $\beta = 23,484$, and $\gamma = 250$. The usual Kolmogorov-Smirnov and Anderson-Darling tests were used to verify the quality of the model, which complies with all the criteria and has thus been identified as the most suitable one. The authors have studied similar models (Marek, Vrabec in [4] for other years and other structure of the data) and can say that, on a long term basis, the three-parametric logistic distribution is the best choice. A detailed study of the evolution of empirical wages in the Czech Republic can be found in [4].

Figure 3 shows the smoothing of the empirical data with the aid of the three-parametric logistic distribution.

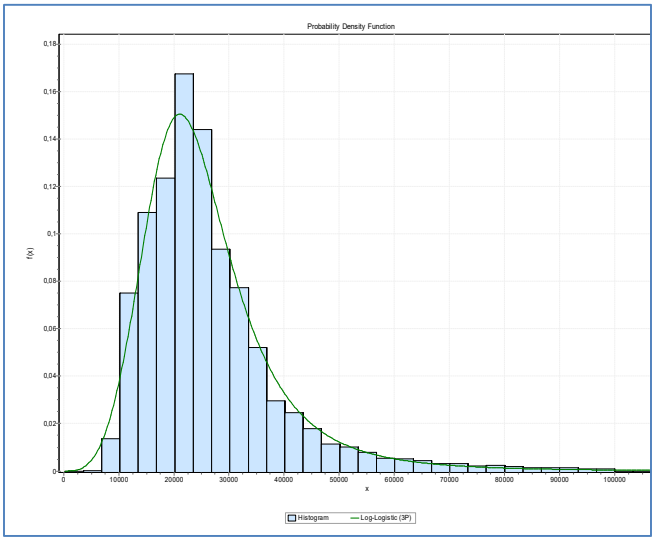


Figure 3 The empirical distribution after smoothing with the aid of the theoretical curve

3 Minimum Wage Scenarios

Increasing the minimum wage is by far not a mere technical issue – setting a lower limit and increasing to this limit the wages of all employees. A natural question arises what will happen with the wages of all the others. Let us take an example: Mr. XY's wage is currently 2,000 CZK higher than the minimum wage. If the minimum wage is increased by exactly 2,000 CZK, he will be paid the minimum wage. But how will Mr. XY feel about such a change? His wage was 2,000 CZK higher than the minimum wage. He will definitely want a raise. But by how much? By the actual amount of 2,000 CZK? Or relatively, to get a wage higher than the minimum wage by the same percentage as before? Other rules of thumb could be considered, of course. Another aspect is the time lag: employees will request raises after some time from the actual increase of the minimum wage. And employers will try to drag out the delay for as long as possible. We have created three scenarios to reflect the impacts of the changes in the minimum wage on the overall wage distribution. We have chosen scenarios which, in our opinion, best correspond to the situation, but other scenarios could be considered as well. Let us now describe the scenarios we have identified.

Scenario 1 (Minimal) is based on an unrealistic assumption that all employees whose wages are below the newly set minimum wage limit will begin to receive this new minimum wage and all the others will simply accept their current wage levels. It can be seen at first sight how unrealistic this assumption is. If an employee's wage was 11,000 CZK and the minimum wage limit was 9 500 CZK their wage was higher by 1 500 CZK (cf. Scenario 2), or nearly 16% higher (cf. Scenario 3) than the minimum wage.

Scenario 2 (Plus) assumes that employees will put up with preserving differences, that is, all wages would be increased by the same amount – in our example, by 1 500 CZK;

Scenario 3 (Multi) is the most realistic in our opinion. It is based on the idea that employees will want to preserve the ratios. In our example, an employee with a wage of 19,000 CZK (twice as much as the minimum wage) will request to get twice as much as the minimum wage increased to 11,000 CZK, , i.e., 22,000 CZK.

Let us now analyze the impacts on the wage distributions in all the described scenarios. Even before studying particular data, we can formulate a few theoretical expectations. It is certain that the location of the wage

distribution will be changed regardless of the scenario. The average value of wages will be increased and the chart will move to the right. For Scenario 1, the variability of wages will be decreased both absolutely and relatively. In Scenario 2, the absolute differences will remain reserved, but on a higher level, and relative variability will thus be decreased. Hence the wages will be more egalitarian in both these scenarios. The skewness will be changed as well: the distribution gets denser on the left-hand side, and the right hand side's skewness is far less affected. The proportions are relatively least affected within Scenario 3, in which the location is also shifted and the absolute variability is increased (even the quickest), but relative variability remains practically unchanged, similar to the other parameters of the distribution's shape. Let us now present the results of simulated calculations.

Scenario 1 – Minimal

If all employees in the categories above the newly set minimum wage (11,000 CZK) accept their existing wages and the only change would be restricted to moving employees from below the limit (there are 66,244 such employees in our sample) to the 11,000-CZK category (in which the original number was 23,047), the number of employees in this category would be increased to 89,291, and the overall average wage would go up from 27,157 CZK to 27,195 CZK; this means an increase in labor costs by 81,176,000 CZK (from 57,557,026,750 CZK to 57,638,202,750 CZK – these costs are only calculated from our sample), and a relative increase would amount to 1.4‰. In our study we take into consideration sensitivity to changes in average wages at 100 CZK plus, the process of the wage increase is terminated immediately. A difference between the theoretical distributions before and after the minimum wage increase would be minimal, cf. Table 1 and Figure 4 below. An overall increase of the labor costs would also be insignificant.

Variable	Value
Sample average	27,157
New average wage	27,196
Minimum wage – current values	9,900
Minimum wage – new value (40% of average)	10,863
Increase in labor costs	0.14%

Table 1 The Minimal Scenario – results

Scenario 2 – Plus

We think this scenario is unrealistic as well – an increase of a wage from 9,900 CZK to 11,000 CZK represents 11.11%, while an increase by the same amount of 1,100 CZK amounts to just a 3.85% for a wage of 28,500 CZK. The time evolution of the minimum and average wages trying to catch each other is illustrated in the Table and Figure below. For the chosen sensitivity of 100 CZK, the situation is stabilized after four steps. The data is shown in Table 2, the movements of the wage distribution is shown in Figure 5. Increase in labor costs is significant, namely, by 9.9%.

Step	New average wage	Minimum wage	Increase in labor costs
1	28,820	11,528	6.1%
2	29,485	11,794	8.6%
3	29,751	11,901	9.6%
4	29,858	11,943	9.9%

Table 2 The Plus Scenario– results

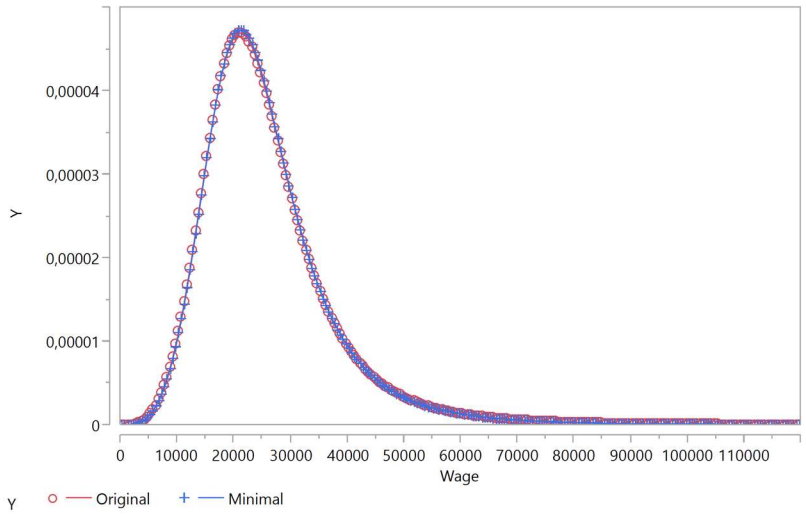


Figure 4 The original wage distribution vs. the Minimal Scenario

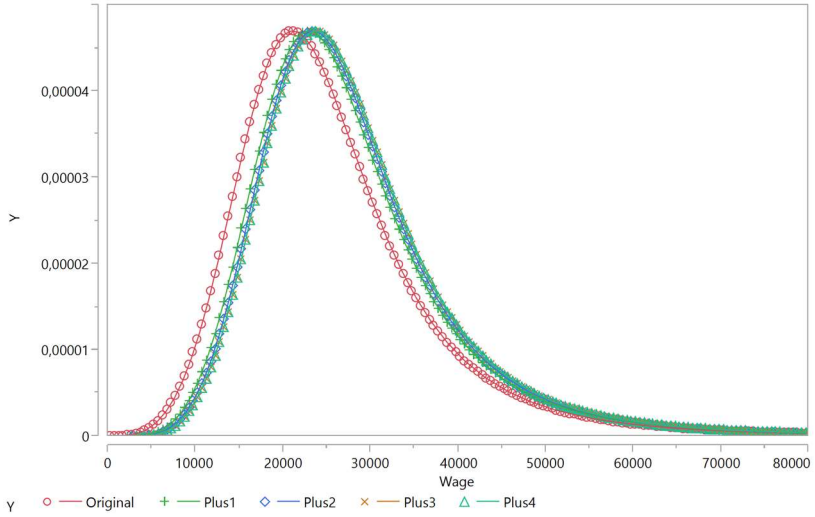


Figure 5 The original wage distribution vs. the Plus Scenario

Scenario 3 – Multi

Regarding employees' behavior, we think this scenario is the most likely one. Due to the above-mentioned time lag and the natural human slackness, the mutual effects of the minimum and average wages will be distributed in time and will not be instantaneous.

Step	New average wage	Minimum wage	Increase in labor costs
1	32,066	12,826	18.1%
2	37,862	15,145	39.4%
3	44,706	17,882	64.6%
4	52,786	21,114	94.4%

Table 3 The Multi Scenario – results

Figure 6 nicely illustrates the movements of the wage distribution. The location is changed (moving to the right), as well as the variability (increasing), the skewness (the positive skewness is growing), and the kurtosis (decreasing). The increase in labor costs is alarming; namely, it amounts to 94.4%!

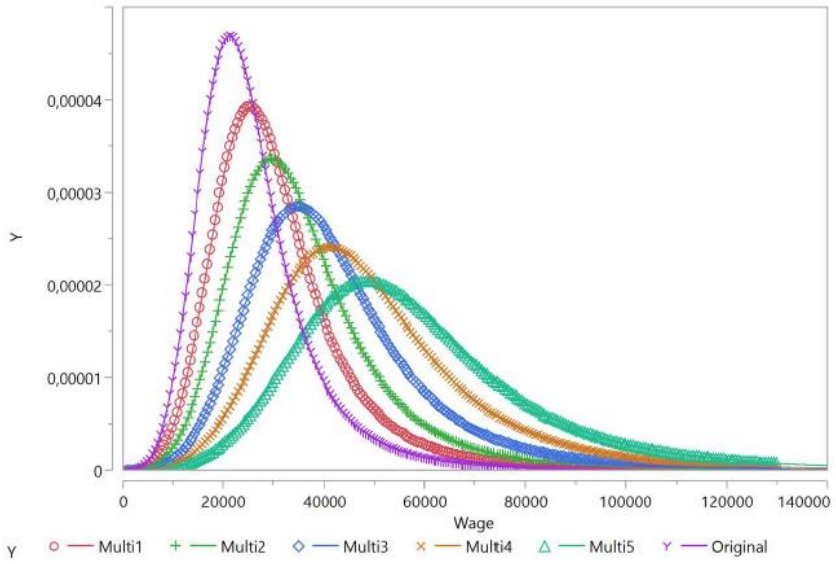


Figure 1 The original wage distribution vs. the Multi Scenario.

4 Conclusions

Scenario 3 appears to be the most realistic from our point of view. Logically, employees will try to preserve the ratio with respect to the minimum wage after each (repeated) increase of the latter. Such response will be neither instantaneous nor automatic; there is a time lag between the changes of the average and minimum wage levels (legislative amendments, tripartite negotiations, employees' attitudes, etc.). This scenario will push the minimum wage up without any limitations by ever increasing the average wage on which the minimum wage is dependent. This cyclic dependence is time-dynamic and the problem has no principal solution. In other words, the minimum wage should not be tied with the average wage and another method for setting out the limit of the minimum wage should be sought for.

Acknowledgements

This paper was written with the support of the Czech Science Foundation project No. P402/12/G097 „DYME – Dynamic Models in Economics“.

References

[1] Aristotle: Metaphysics. Retrieved from http://classics.mit.edu/Aristotle/metaphysics.html.
[2] Bartosová, J. and Longford, N. T.: A study of income stability in the Czech Republic by finite mixtures, *Prague Economic Papers* **23** (3), (2014), 330–348.
[3] Malá, I.: Multidimensional probabilistic model for Czech household income distribution, *Politická ekonomie* **63** (7), (2015), 895–908. (In Czech).
[4] Marek, L.: Analysis of wage evolution in the Czech Republic from 1995 to 2008, *Politická ekonomie* **58** (2), (2010), 186–206. (In Czech).
[5] Marek, L., Vrabec, M.: Using mixture density functions for modelling of wage distributions. *Central European Journal of Operations Research* **24** (4), (2016), 389–405. doi: 10.1007/ s10100-015-0409-4.

NEG Methodology in Socio-Economic Development and EU Labour Market Research

Adrianna Mastalerz-Kodzis [1], Ewa Pośpiech [2]

Abstract. The search for the causes of an uneven socio-economic development of countries is one of the main goals of NEG. Economic activities are not evenly spread out in space. It is possible to indicate clusters of population, wealth as well as poverty. Causal relationship analysis, as well as the analysis of spatial relationship between economic characteristics forms the key to recognizing contemporary economy. In the article, we have determined, on the base of the data considering selected economic characteristics of EU countries, the matrix of distance between countries, we have calculated the HDI indicator, the Gini coefficient, the quotient of income and population' potentials, and we have studied the relationship between the obtained values. The article focuses on the socio-economic and linked to the EU countries labour market characteristics. The article is composed of two basic parts: a short discussion on analyses' methodology and the results of empirical research.

Keywords: socio-economic development, HDI coefficient, Gini coefficient, new economic geography, the quotient of potentials method.

JEL Classification: C33, C44
AMS Classification: 91B40, 91B45

1 Introduction

The foundation of the European Union (EU) and its rules of functioning significantly influence the socio-economic development of member countries. A free flow of population as well as the liberty in employment undertaking positively influence most of the economic indicators, promote the development of trade, open the boarders for the migration of population and allow for a free development of economic activity.

The economic development of countries depends, among others, on their geographical location. It can be shown, that there exists a considerable influence of the surrounding on the economic development level of a given territorial unit, as well as on the labour market development of this unit. The aim of the article is to illustrate the extent of this phenomenon using selected economic characteristics for EU countries.

NEG (New Economic Geography) deals with spatial location of economic activity. Considered are, for example, the results of the globalization phenomenon, the development of agglomeration and metropolis. It can be proved that spatial location (geographic co-ordinates, distance, neighbourhood) considerably influences the development and formation of a given phenomenon or economic process. Methods and models of NEG are used in socio-economic analyses of spatial character, inter alia, in the study of economic development of agglomeration, localization of industry, trade, globalization. There exists a possibility of studying the relationship between territorial units. The models include not only the own potential of a unit but also the spatial location of a region and the influence of the neighbouring regions.

A commonly used measure of the socio-economic development level is the HDI indicator. However, analysing the development level we should also consider the equability of income arrangement as well as the level of labour market development, for these are significant determinants of the population's life standard. In the article, we have determined, on the base of the data considering selected economic characteristics of EU countries, the matrix of distance between the capitals of EU countries, we have calculated the Gini coefficient, the HDI indicator and the quotient of income and population's potentials. We have studied the relationship between the obtained values.

[1] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, 1 Maja 50, 40-287 Katowice, adrianna.mastalerz-kodzis@ue.katowice.pl
[2] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, 1 Maja 50, 40-287 Katowice, ewa.pospiech@ue.katowice.pl

2 Elements of research methodology

In the empirical analyses we used selected statistics and econometrics methods. We used, among others, the works of [2], [3], [12], [14].

2.1 The method of potentials and the quotient of potentials

In the year 1970 G. Dutton introduced the concept of income and population quotient of potentials. This quotient is the equivalent of income per person, it includes, however, the relations between regions which form these potentials. The formulas below were constructed on the base of the work of [5], [14], they were, however, modified without doubling own potentials. In the work of [4] the income potential is treated as a measure of the accessibility of economic activity. Keeble claims that there exists a significant spatial correlation between the variability of population's potential and the variability of many indicators proclaiming the level of socio-economic development of territorial units. Therefore, according to Kebble, population's potential can be treated as a substitute of socio-economic phenomena.

Let there be given a division system of a certain territory consisting of n spatial units $(n \in N)$ (e.g. countries). In i - this country we accept the symbols: L_i - population's potential, Z_i - income potential, P_i - quotient of potentials, z_i - GDP, l_i - number of people, g_i - GDP per capita, d_{ij} - distance between territorial units i and j; we accept that $d_{ii} = 1$ (in this article, as d_{ij} we have accepted the Euclidean distance between EU countries' capitals, we have used geographical co-ordinates).

We define units' potentials as the sums of own and surrounding's potentials:

$$L_i = \frac{l_i}{d_{ii}} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{l_j}{d_{ij}} \tag{1}$$

$$Z_i = \frac{z_i}{d_{ii}} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{z_j}{d_{ij}} \tag{2}$$

and the quotient of potentials i:

$$P_i = Z_i / L_i \tag{3}$$

Potential determines the force of the effect of all territorial units on a given unit. It is the total of two components: own potential and surrounding's potential. A territorial unit may possess a small own potential, but it can be located in a favourable geographic site, nearby strong units, then its potential increases.

Potential is also understood as a measure of spatial accessibility of a territorial unit, it determines the intensity of interaction between units taking into consideration economic values and the distance between units. Population's potential (1) means the accessibility of a given country to the population of the other countries taking into consideration the metrics of distance. Income potential (2) is an income function created in a certain territorial unit and in other units with the consideration of the distance between them.

The value of the quotient P_i / g_i depends on the proportion between the potential of a territorial unit surrounding and own potential of this unit in the range of income, population number and distance:

- If $P_i / g_i > 1$, then the unit potential is bigger than income per inhabitant; the surrounding significantly influences the forming of economic growth level. In such situation are the countries with small own potential, located on a favourable geographic site, nearby strong countries, with high own potential.
- When $P_i / g_i < 1$, then the income per inhabitant is higher than the unit's potential. Such a situation takes place in the case of economically well developed units.

2.2 Index HDI construction

The HDI (Human Development Index) indicator is used to determine the level of development of the countries in the world ([1], [6], [7], [8], [13]). In the evaluation process, the synthetic indicator HDI takes into account three criteria: a long and healthy life, the level of education and life standard. The US programme concerning development recommends calculating the HDI indicator with the use of basic characteristics which are:

- **Heath Index**, for LE_i - average lifespan in i- this country according to the formula:

$$H.Ind_i = \frac{LE_i - 20}{65} \tag{4}$$

- **Education Index**, for LIT_i - expected years of schooling and ENR_i - mean years of schooling according to the formula:

$$E.Ind_i = \frac{(LIT_i/18)+(ENR_i/15)}{2} \tag{5}$$

- **Welfare Index**, where y_i is the income per capita ($) in a given country:

$$(Y.Ind_i) = \frac{\log(y_i)-\log(\$100)}{\log(\$75)-\log(\$100)} \tag{6}$$

- **The social development index** per capita for a given country is calculated according to the formula:

$$HDI_i = \sqrt[3]{H.Ind_i \cdot E.Ind_i \cdot Y.Ind_i} \tag{7}$$

It is accepted that the value of HDI on a given level indicates a country which is: very high developed (0,901-1), high development (0,801-0,9), medium developed (0,501-0,8), poorly developed (0-0,5).

2.3 Gini coefficient

A commonly used measure of income inequalities is the Gini coefficient, which, in econometrics, is called the social inequality index. In economy, the Gini coefficient is primarily used in the measurement of an uneven location of goods, mainly in case of an uneven location of income of e.g. households. The value of the coefficient belongs to the range from 0 to 1. In the case when the value equals 0, it means that all the people receive the same amount of income. When the value of the coefficient reaches the level of 1, it means that the whole of the income of a given country is concentrated in one household. The reading of Gini coefficient should be interpreted as follows: the higher the level of Gini coefficient, the greater the inequalities in income in a given country.

When the observations are sorted ascending, then it is calculated using the formula ([11]):

$$G = \frac{\sum_{i=1}^{n}(2i - n - 1)z_i}{n^2\bar{z}} \tag{8}$$

where n is the number of sample items, z_i is the income of i-this unit, $\bar{z}$ is the average income in the sample. We can also gives an alternative formula which is:

$$G = \frac{E|X - Y|}{2\mu} \tag{9}$$

for $E|X - Y|$ - average absolute difference between two incomes, μ is the average value of incomes. With a doubled value of G coefficient we obtain the information about what part of the mean μ is the average absolute difference between a random pair of incomes (e.g. for G = 0.3 the average difference between the incomes makes up 60% of average income.).

3 Selected empirical study results

Empirical analysis was conducted on the base of selected socio-economic characteristics for the EU countries in 2015. Using geographical co-ordinates we measured the matrix of the distance between the EU countries' capitals. We used the metrics of Euclidean distance. We gathered the following data: GDP, population number, average lifespan, expected years of schooling, mean years of schooling (see also: [9], [10]). We determined the income potential and its component of surrounding, the population potential and its component of surrounding, as well as the quotient of potentials (Table 1, Figure 1).

Analyzing the values presented in the table, we can state that :

- The highest values of income potential belong to: Germany, United Kingdom, France, Italy, Belgium, Czech Republic; where the countries with the highest own potential are: Germany, United Kingdom and France,

and the countries with the highest surrounding's potential: Belgium, Luxembourg, Czech Republic and Slovakia.

- The lowest values of income potential belong to Cyprus, Portugal, Malta, Finland, where the countries with the lowest own potential are: Cyprus, Estonia, Malta, and the countries with the lowest surrounding's potential: Cyprus and Greece.
- The highest values of population's potential belong to: Germany, United Kingdom, France, Italy, Czech Republic, Poland; where the countries with the highest own potential are: Germany, United Kingdom, France and Italy, and the countries with the highest surrounding's potential; Belgium, Luxembourg, Czech Republic and Slovakia.
- The lowest values of population's potential belong to: Cyprus, Malta, Finland, where the countries with the lowest own potential are Luxembourg and Malta, and the country with the lowest surrounding's potential is Cyprus.

EU countries	Z_i	$\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{z_j}{d_{ij}}$	z_i	L_i	$\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{l_j}{d_{ij}}$	l_i	P_i
Netherlands	3103839	2320118	783721	78,70	61,80	16,90	39440
Germany	4967782	1412705	3555077	122,32	41,62	80,70	40612
Demark	2495896	2248431	247466	64,25	58,55	5,70	38846
Ireland	1559864	1315939	243925	40,40	35,70	4,70	38613
Sweden	1741380	1297480	443901	46,19	36,39	9,80	37702
United Kingdom	4319342	1818169	2501173	113,38	48,68	64,70	38095
France	4404044	2001473	2402571	117,35	52,95	64,40	37528
Austria	2475209	2110897	364312	69,34	61,04	8,30	35695
Belgium	3308978	2844118	464859	85,48	74,18	11,30	38711
Luxembourg	2583757	2527625	56132	67,53	66,93	0,60	38262
Finland	1265628	1053092	212537	36,14	30,64	5,50	35022
Slovenia	2193208	2132430	60778	62,80	60,70	2,10	34923
Italy	3253394	1244891	2008503	94,75	34,95	59,80	34335
Spain	2592345	1079620	1512725	75,75	29,65	46,10	34222
Czech Republic	3112788	2799836	312953	83,44	72,94	10,50	37305
Greece	1247327	976540	270787	40,28	29,28	11,00	30965
Cyprus	702232	665860	36372	21,09	19,89	1,20	33298
Estonia	1352713	1317704	35009	38,59	37,29	1,30	35055
Lithuania	1322632	1246081	76551	40,28	37,38	2,90	32837
Poland	2338093	1379423	958670	76,52	37,92	38,60	30556
Slovakia	2442954	2385427	57527	68,86	66,76	2,10	35478
Malta	1222948	1211420	11529	35,48	35,08	0,40	34464
Portugal	1291799	1016892	274907	38,71	28,41	10,30	33368
Hungary	1963051	1720758	242293	59,77	49,87	9,90	32843
Croatia	2089913	2004107	85806	61,09	56,89	4,20	34208
Latvia	1317182	1271926	45256	39,25	37,25	2,00	33561
Romania	1437581	1049024	388557	50,69	31,19	19,50	28361
Bulgaria	1373121	1252733	120388	45,59	38,49	7,10	30120

Table 1 Income and population potential (own and of the surrounding) and the quotient of potentials for EU countries in 2015

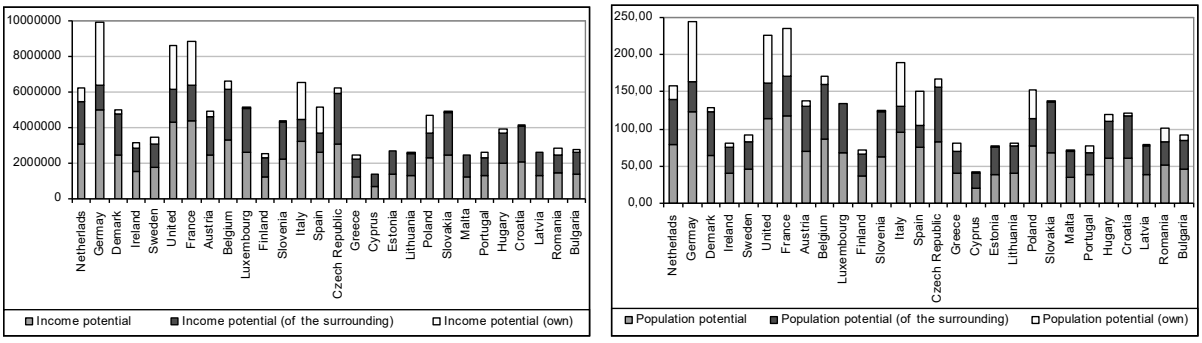


Figure 1 Income and population's potential and their components for EU countries in 2015

Next, we determined the HDI indicator and the Gini coefficient for the EU countries and compared their values with the P_i / g_i quotient. The results are presented in Table 2 and Figure 2.

Population's potential determines the accessibility of a territorial unit to the population of the other units in terms of distance. Income potential of a unit is a function of the income created in this unit and in other units as well as of the distance between units. It takes into consideration the influence of income dependence on spatial income variability. Analyzing the results, we can state that:

- The socio-economic development of EU countries measured with HDI coefficient is on a very high level (27 out of 28 countries possess HDI > 0.9),
- The Gini indicator belongs to the range (0.237; 0.379). Countries with the lowest level of income inequalities are: Slovakia, Slovenia, Czech Republic and Finland, whereas the highest level of inequalities was noted in: Lithuania, Romania and Bulgaria.
- P_i / g_i value depends on the proportion between potential generated by the surrounding and own potential in terms of income as well as in terms of population number and distance.
- For 10 countries (the wealthiest ones) $P_i < g_i$ relation takes place. These are the countries with the highest P_i and g_i values. They possess a higher spatial concentration degree of income potential than population's potential. When P_i / g_i <1, then income per inhabitant exceeds the region's potential.
- For the other 18 countries $P_i > g_i$ relation takes place. These are mostly countries with lower values of both P_i, as well as g_i. They possess a lower spatial concentration degree of income potential than population's potential. It can be stated that the country's surrounding plays a greater part in generating income potential than in generating its population's potential. When P_i / g_i >1 it means that the population's potential of a country is higher than its income per inhabitant. We can, therefore, point out the importance of the surrounding in forming of the economic growth development level. The highest values of P_i / g_i belong to countries with a low level of income.

EU countries	GINI	HDI	P_i/g_i		EU countries	GINI	HDI	P_i/g_i
Netherlands	0,267	0,922	0,850		Czech Republic	0,250	0,870	1,252
Germany	0,301	0,916	0,922		Greece	0,342	0,865	1,258
Demark	0,274	0,923	0,895		Cyprus	0,336	0,850	1,099
Ireland	0,298	0,916	0,744		Estonia	0,348	0,861	1,302
Sweden	0,252	0,907	0,832		Lithuania	0,379	0,839	1,244
United Kingdom	0,324	0,907	0,985		Poland	0,306	0,843	1,230
France	0,292	0,888	1,006		Slovakia	0,237	0,844	1,295
Austria	0,272	0,885	0,813		Malta	0,281	0,839	1,196
Belgium	0,262	0,890	0,941		Portugal	0,340	0,830	1,250
Luxembourg	0,285	0,892	0,409		Hungary	0,282	0,828	1,342
Finland	0,252	0,883	0,906		Croatia	0,306	0,818	1,674
Slovenia	0,245	0,880	1,207		Latvia	0,354	0,819	1,483
Italy	0,324	0,873	1,022		Romania	0,374	0,798	1,423
Spain	0,346	0,876	1,043		Bulgaria	0,370	0,782	1,776

Table 2 The values of HDI and Gini coefficients and P_i/g_i for EU countries in 2015

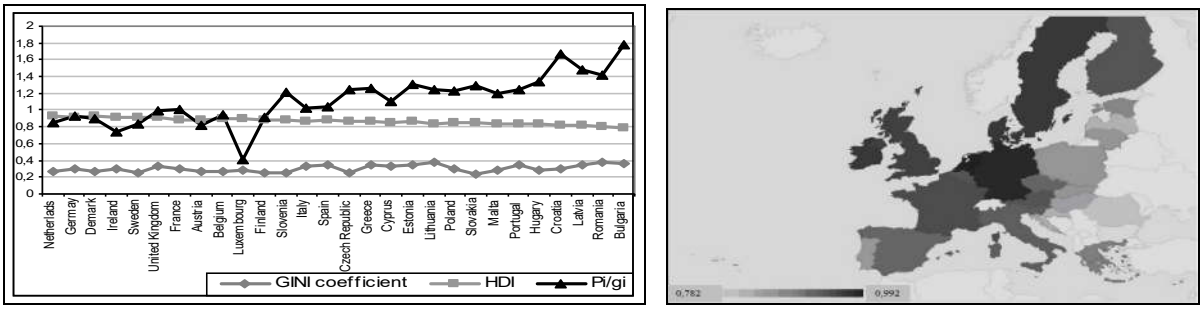


Figure 2 The values of: HDI, Gini coefficients, P_i/g_i and the map of HDI values for EU countries in 2015

We determined Pearson's correlation coefficients between measures. It can be stated that:

- There exists a significant, negative relationship between the HDI coefficient and the value of P_i/g_i (the value of Pearson's coefficient -0.85). Then, the higher the level of socio-economic development (and at the same time also a higher value of GPD), the lower the influence of the surrounding on a given country. The lower the level of socio-economic development (poorer countries), the higher the influence of the surrounding.
- A moderate, negative relationship between the HDI and the Gini coefficient was observed. (Pearson -0.53). Therefore, the higher the development level of a given country, the smaller the income inequalities. The highest level of income inequalities belongs to countries with the lowest HDI.
- There exists a moderate, positive relationship between the HDI and GDP per capita (Pearson 0.67). GPD per capita is one of the three components of HDI (Pearson 0,67), it explains around 46% of HDI variations. The health and education components are responsible for the rest.
- A moderate, positive relationship was noted between the value of the Gini coefficient and the value of P_i/g_i (Pearson 0,45). The higher the level of income inequalities, the more dependent the country is on the surrounding.

- There exists a low, negative relationship between GDP per capita and the Gini coefficient (Pearson -0.38). In approximately 15% of the cases, the income inequalities can be explained by the levels of income in EU countries.

All analyzed indicators: HDI, Gini coefficient and P_i/g_i values influence population's life standard. In the countries with a high level of GDP per capita, a high index of socio-economic development, a relatively low level of the Gini coefficient, population's life standard is considerably higher than in less developed countries. The labour market development level determines all mentioned in the article indicators to a large extent. The problem of many EU countries is a high level of unemployment (Spain, Greece), especially among young people (Spain, Greece, Croatia, Italy). It is the EU authorities' duty to optimally make use of the population's potential of these countries, what seems to be highly significant, for, according to Keeble and others, population's potential can be treated as a substitute of socio-economic phenomena. Therefore, in order to fully describe the socio-economic situation of countries, we should consider not only the indicators which state the level of income, education and health, but we should also take into consideration the differentiation of income values (income inequalities), economic surrounding as well as labour market level.

4 Conclusion

The socio-economic development of EU countries is differentiated both in terms of health and education level as well as income and income inequalities. There are countries very highly developed, with a high GDP level and a low income inequalities' level, where own potential of income and population is dominant over surrounding's potential. We can also indicate poorer countries, less developed, where the differentiation considering income is significant, whereas the surrounding plays an important part in the process of economic life development. The obtained results of study show that in pursuit of reducing the inequalities between EU countries a special part should be assigned to the influence that comes from large countries. In the era of globalization, geographical location of e.g. a newly formed economic activity or the choice of the place of living of a EU citizen should be preceded by an analysis of selected socio-economic characteristics.

References

[1] Biernacki, M.: Kilka uwag o pomiarze dobrobytu społecznego. *Mathematical Economics* **3(10)** (2006), 115-126.
[2] Getis A,., Mur, J., and Zoller, H.: *Spatial Econometrics and Spatial Statistics*, Palgrave Macmillan, New York, 2004.
[3] Hsiao, C.: *Analysis of Panel Data*, Cambridge University Press, 2003.
[4] Keeble, D., Owens, P.L., and Thompson, C.: Regional Accessibility and Economic Potential in the European Community, *Regional Studies* **6(16)** (1982), 419-432.
[5] Krugman, P.R.: *Geography and Trade*, The MIT Press, Cambridge, Massachusetts, 1991.
[6] Mastalerz-Kodzis, A., and Pośpiech, E.: Dynamic and Spatial Analysis of Economic and Labour Market Characteristics in European Countries. In: *Mathematical Methods in Economics*, Liberec, (2016A), 546-551.
[7] Mastalerz-Kodzis, A., and Pośpiech, E.: Economic Development of European Countries in Globalization Conditions – Quantitative Analysis. In: *Globalization and its socio-economic consequences.* 16th International Scientific Conference Proceedings, Part III, Zilina (2016B), 1791-1798.
[8] Mastalerz-Kodzis, A.: Dynamika przemian społeczno-ekonomicznych krajów Unii Europejskiej. *Studia ekonomiczne*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach **265** (2016), 26-37.
[9] Mastalerz-Kodzis, A., and Pośpiech, E.: Wielowymiarowa analiza porównawcza w ujęciu dynamicznym na przykładzie wybranych charakterystyk ekonomicznych. In: *Metody ilościowe w badaniach ekonomicznych*, **16(4)** (2015), 24-33.
[10] Mastalerz-Kodzis, A., and Pośpiech, E.: Spatial and Temporal Analysis of Labour Market Characteristics. *Folia Oeconomica Stetinensia* **16(2)** (2016), 60-71.
[11] Panek, T.: *Statystyka społeczna*, PWN, Warszawa, 2014.
[12] Rencher, A. C.: *Methods of Multivariate Analysis*, USA, John Wiley & Sons, 2002.
[13] Stanton E. A.: *The Human Development Index: A History*. Political Economy Research Institute PERI **127**, 2007.
[14] Suchecki, B. (ed.): *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*. C.H. Beck, Warszawa, 2010.
Data: UNDP Human Development Report, Eurostat, (22.04.2017).

Early Exercise Premium and Boundary in American Option Pricing Problem

Kateřina Mičudová[1], Ladislav Lukáš[2]

Abstract. The paper is focused on American option pricing problem. Assuming non-dividend paying American put option leads to two disjunctive regions, a continuation one and a stopping one, which are separated by an early exercise boundary. We present variational formulation of American option problem with special attention to early exercise action effect. Next, we discuss financially motivated additive decomposition of American option price into a European option price and another part due to the extra premium required by early exercising the option contract. As the optimal exercise boundary is a free boundary, its determination is coupled with the determination of the option price. Therefore, a closed-form expression of the free boundary is not attainable in general. We discuss in detail a derivation of an asymptotic expression of the early exercise boundary. Finally, we present some numerical results of determination of free boundary based upon this approach. All computations are performed by sw Mathematica, and suitable numerical procedure is discussed in detail, as well.

Keywords: American option, early exercise premium, early exercise boundary, pricing problem.

JEL classification: G13
AMS classification: 91G80

1 Introduction

American option pricing problems given with different levels of abstraction are presented in [3], [4], [6], and [7]. In the paper, we present variational formulation of American option problem with special attention to early exercise action effect, first. Next, we discuss financially motivated additive decomposition of American option price into two parts, i.e. the European option price and the extra premium required by early exercising the option contract, in particular. Finally, our main contribution is presented within the third section which is focused mainly upon numerical solution of optimal exercise bounary detection problem. We know that in contrast to European options, American options can be exercised at any time between the writing and the expiration of the option contract. Exercising flexibility gives American option more profiteering opportunity than European option offers, and it means the payoff functions of American options take the following general form, where an index C, and P denotes an option type, i.e. *call*, and *put*, respectively

$$\begin{aligned} V_C(S,t) &\geq \max(S-K,0),\ \forall (S,t) \in \mathbb{R}_+ \times (0,T], \quad \text{for call option}, \\ V_P(S,t) &\geq \max(K-S,0),\ \forall (S,t) \in \mathbb{R}_+ \times (0,T], \quad \text{for put option}. \end{aligned} \tag{1}$$

To be specific, let us take non-dividend paying American put option. Thorough inspection of relations (1) gives an idea that for American put option there exist two disjunctive subregions, say Σ_1 and Σ_2, respectively, covering $\mathbb{R}_+ \times [0,T]$, i.e. $\Sigma_1 \cup \Sigma_2 = \Sigma = \mathbb{R}_+ \times [0,T]$. Let $\partial\Sigma_1, \partial\Sigma_2$ denote boundaries of Σ_1, Σ_2, defined by $\partial\Sigma_i = \mathrm{cl}(\Sigma_i)\backslash\Sigma_i, i = 1,2$, where $\mathrm{cl}(\Sigma_i)$ is a closure of set Σ_i. Hence, a common piece of their boundaries, denoted $\Gamma = \partial\Sigma_1 \cap \partial\Sigma_2$, is called optimal exercise boundary. The optimal exercise boundary is reasonably defined by mapping $\Gamma : [0,T] \to \Gamma(t)$ thus specifying the asset price at Γ, denoted $S_\eta = \Gamma(t)$, in particular, where the subscript η instead of usual t is to point to specific underlying asset price at the optimal exercise time t_e. The optimal exercise boundary Γ is not given a priori and has to be determined together with option pricing function $V(S,t)$ by solving the corresponding option pricing problem. These subregions are defined as follows

$$\begin{aligned} V(S,t) &> \max(K-S,0),\ \forall (S,t) \in \Sigma_1, \\ \Sigma_1 &= \{(S,t)\,|\,\Gamma(t) \leq S < +\infty\},\ \forall t \in [0,T]\}, \end{aligned} \tag{2}$$

$$\begin{aligned} V(S,t) &= \max(K-S,0),\ \forall (S,t) \in \Sigma_2, \\ \Sigma_2 &= \{(S,t)\,|\,0 \leq S \leq \Gamma(t)\},\ \forall t \in [0,T]\}, \quad \Gamma(t) < K, \forall t \in [0,T), \end{aligned} \tag{3}$$

[1]University of West Bohemia, Faculty of Economics, Dept. of Economics and Quantitative Methods, Univerzitní 22, 306 14 Pilsen, pitrovak@kem.zcu.cz
[2]University of West Bohemia, Faculty of Economics, Dept. of Economics and Quantitative Methods, Univerzitní 22, 306 14 Pilsen, lukasl@kem.zcu.cz

where Σ_1 is called the continuation subregion, since the payoff is zero when $S > S_\eta$, and the holder should continue to keep the option, whereas Σ_2 is called stopping subregion. The asset prices S_η on the exercise boundary are denoted $S_\eta = \Gamma(t)$ precisely, as we have already stated above. Noting, the continuation subregion Σ_1 is sometimes called retained subregion alternatively, whereas the stopping subregion Σ_2 is called selling subregion, or exercise subregion, as well. In Figure 1, there are depicted both subregions Σ_1, Σ_2, encountered within an American put option pricing problem, schematically, where the horizontal axis represents an underlying asset price S, and vertical axis stands for time t elapsed from entering into option contract, showing the expiration date T, and exercise price K, too. Following classical procedure based on construction of self-financing portfolio,

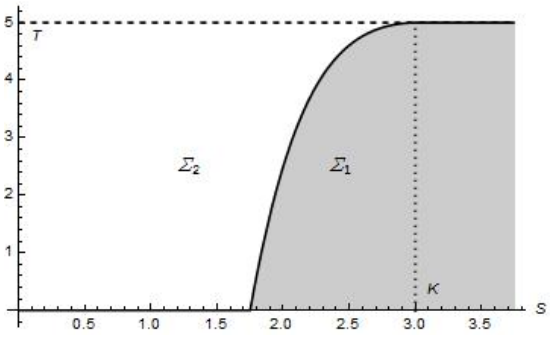


Figure 1: Early exercise boundary

Δ-hedging principle, and Itô's formula, which is common and typical for derivation of almost all models within the field of financial option pricing problems, we can infer that $V(S,t)$ satisfies the Black-Scholes PDE in the subregion Σ_1

$$\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV = 0,\ \forall (S,t) \in \Sigma_1, \tag{4}$$

assuming $V(S,t) \in C^{2,1}(\Sigma_1)$, in classical formulation framework.

The boundary conditions on $\partial\Sigma_1$ are specified in following way. First, on the optimal exercise boundary Γ, it holds

$$V(S,t) = \max(K-S,0) \ \Rightarrow\ \frac{\partial V(S,t)}{\partial S} = -1,\ S = S_\eta \in \Gamma(t),\ \forall t \in [0,T). \tag{5}$$

Further, the terminal condition at $t = T$, and the condition when $S \to +\infty$, are following

$$V(S_T,T) = \max(K-S_T,0), \quad \lim_{S\to+\infty} V(S,t) = 0,\ \forall t \in [0,T). \tag{6}$$

Concluding, we see that American put option pricing problem leads to solving a function pair $\{V(S,t), \Gamma(t)\}$ in subregion Σ_1 satisfying PDE (4) and boundary-terminal conditions (5) and (6).

Following [7], we are able to get more compact form of the problem and to clarify its setting, too. First, it is reasonable to introduce so called Black-Scholes differential operator $\mathcal{L}_{\mathrm{BS}}$ as follows

$$\mathcal{L}_{\mathrm{BS}} G(S,t) = \frac{\partial G}{\partial t} + rS\frac{\partial G}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 G}{\partial S^2} - rG, \quad \forall (S,t) \in \Sigma, \quad \Sigma = \mathbb{R}_+ \times (0,T), \tag{7}$$

which is defined by its application on any function $G(S,t) \in C^{2,1}(\Sigma)$.

Using (2), (3) and $\mathcal{L}_{\mathrm{BS}}$, we can rewrite properties of function $V(S,t)$ in subregions Σ_1, and Σ_2 in compact form

$$\mathcal{L}_{\mathrm{BS}} V(S,t) = 0, \quad V(S,t) > \max(K-S,0), \quad \forall (S,t) \in \Sigma_1. \tag{8}$$

$$\mathcal{L}_{\mathrm{BS}} V(S,t) = -rK < 0, \quad V(S,t) = \max(K-S,0), \quad \forall (S,t) \in \Sigma_2. \tag{9}$$

Since (9) holds on Γ, being tackled as a part of $\partial\Sigma_2$, now, we can conclude by direct computation that both the pricing function $V(S,t)$, and its first derivative with respect to S, i.e. $\frac{\partial V}{\partial S}$, are continuous on the optimal exercise boundary Γ. Combining (8) with (9), we can formulate the American put option problem on $\Sigma = \Sigma_1 \cup \Sigma_2 = \mathbb{R}_+ \times [0,T]$, which is called variational formulation of American put option problem in strong sense, see [4], Chapter 6.2

$$\begin{gathered} \min\{-\mathcal{L}_{\mathrm{BS}} V(S,t),\ V(S,t) - \max(K-S,0)\} = 0, \quad \forall (S,t) \in \Sigma \\ V(S_T,T) = \max(K-S_T,0),\ S_T \in \mathbb{R}_+, \quad V(S,t) \to 0,\ S \to +\infty,\ \forall t \in [0,T]. \end{gathered} \tag{10}$$

Following [5] and [7], the variational formulation of American put option pricing problem in weak sense is

$$\begin{aligned} &\text{Find } u \in L^2(0,T;W), \ \text{ with } \frac{\partial u}{\partial t} \in L^2(0,T;W'), \\ &\text{and with } \quad u(S,0) = u_0(S),\ \forall S \in \mathbb{R}_+, \ \ u(S,t) \in \mathcal{Q},\ \forall S \in \mathbb{R}_+,\ a.e.t \in (0,T), \\ &\text{so that } \ \Big(\frac{\partial u(t)}{\partial t}, w-u\Big) + a_t(u(t), w-u) \geq 0,\ \forall w \in \mathcal{Q},\ \text{a.e. } t \in (0,T), \end{aligned} \tag{11}$$

where $\mathcal{Q}$ denotes a set of feasible solutions of American put option problem in weak sense which is a closed convex set $\mathcal{Q} \subset W$, assuming the boundary value function $u_0(S)$ being understood in sense of traces

$$\mathcal{Q} = \{v \in W \,|\, v \geq u_0(S), \forall S \in \mathbb{R}_+\}. \tag{12}$$

2 Decomposition formula

Early exercise premium is discussed from several viewpoints in [1], and [2]. Following [4], we present a decomposition formula for American put option $V_P(S,t)$ which gives a link to European put option

$$V_P(S,t) = V_P^E(S,t) + e(S,t), \tag{13}$$

where $V_P^E(S,t)$ is the well-known European put option price on the same underlying asset, and $e(S,t)$ is the early exercise premium given by (14) in following way

$$e(S,t) = Kr \int_t^T \mathrm{d}\eta \int_0^{S_\eta} G(S,t;\xi,\eta)\mathrm{d}\xi, \tag{14}$$

where $G(S,t;\xi,\eta)$ is the fundamental solution, sometimes called Green function, too, of the B-S PDE. Further, the upper limit S_η of the inner integral shows that early exercise premium depends upon the position of the optimal exercise boundary cleary.

The fundamental solution of B-S PDE is defined as a solution of terminal value problem for classical Black-Scholes PDE with a very special terminal condition using the Dirac delta function $\delta(.)$, which is well-known by its properties: $\delta(x) = +\infty, x = 0$, otherwise, $= 0, x \neq 0$, and $\int_{-\infty}^{+\infty} \delta(x)\mathrm{d}x = 1$. Determination of $G(S,t;\xi,\eta)$ starts precisely by solution of the B-S PDE with the special terminal condition

$$V(S,T) = \delta(S-\xi), \quad \xi \in \mathbb{R}_+, \tag{15}$$

thus providing the solution $V(S,t;\xi)$, which directly yields the solution, denoted $G(S,t;\xi,T)$ and given by (16) explicitly

$$G(S,t;\xi,T) = \frac{e^{-r(T-t)}}{\xi\sigma\sqrt{2\pi(T-t)}} \exp\Big[-\frac{1}{2\sigma^2(T-t)}\Big(\ln\frac{S}{\xi} + (r-\frac{\sigma^2}{2})(T-t)\Big)^2\Big]. \tag{16}$$

Tackling $G(S,t;\xi,\eta)$ as a function of ξ,η, we know that it is the solution of the so-called adjoint Black-Scholes PDE. Let $v(\xi,\eta) = G(S,t;\xi,\eta)$, then function $v(\xi,\eta)$ satisfies an adjoint terminal value problem

$$\begin{aligned} -\frac{\partial v}{\partial \eta} - r\frac{\partial(\xi v)}{\partial \xi} + \frac{1}{2}\sigma^2\frac{\partial^2(\xi^2 v)}{\partial \xi^2} - rv = 0, \\ v(\xi,\eta) = \delta(\xi - S), \\ 0 < \xi < +\infty,\ 0 < S < +\infty,\ t < \eta. \end{aligned} \tag{17}$$

For more details, we refer to [4], Chapter 6.3, and Theorem 6.3 and 6.4 therein.

3 Early exercise boundary

As the optimal exercise boundary Γ is not known a priori, but has to be determined together with valuation function $V(S,t)$, therefore the problem is called *free-boundary value problem* for parabolic PDE. Looking at it, we can recognize the problem to be already formulated in mathematically elegant way. However, we have to point out that the major difficulty under this setting is that one needs to solve for function $V(S,t)$ along with the unknown optimal exercise boundary $\Gamma(t)$, simultaneously.

An asymptotic expression of early exercise boundary near $t = T$ is given in [8]. Following [4], Chapter 6.3, Theorem 6.5, we are able to write an equation to determine the optimal exercise boundary $\Gamma(t)$ in following way

$$\Gamma(t) = K + \Gamma(t)e^{-q(T-t)}\Phi(-\frac{-\ln\frac{\Gamma(t)}{K} + \beta_2(T-t)}{\sigma\sqrt{T-t}}) - Ke^{-r(T-t)}\Phi(-\frac{-\ln\frac{\Gamma(t)}{K} + \beta_1(T-t)}{\sigma\sqrt{T-t}})$$
$$- Kr\int_t^T e^{-r(\eta-t)}\big(1-\Phi(\frac{\ln\frac{\Gamma(t)}{\Gamma(\eta)} + \beta_1(\eta-t)}{\sigma\sqrt{\eta-t}})\big)\mathrm{d}\eta + q\,\Gamma(t)\int_t^T e^{-q(\eta-t)}\big(1-\Phi(\frac{\ln\frac{\Gamma(t)}{\Gamma(\eta)} + \beta_2(\eta-t)}{\sigma\sqrt{\eta-t}})\big)\mathrm{d}\eta, \tag{18}$$

where q is dividend rate, $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{\xi^2}{2}}\mathrm{d}\xi$ is cumulative distribution function of $N(0,1)$, and

$$\beta_1 = r - q - \frac{\sigma^2}{2}, \quad \beta_2 = r - q + \frac{\sigma^2}{2}.$$

However, it is a nonlinear Volterra integral equation of the second kind. Generally, solving such type of equation, there is a rather challenging problem for numerical mathematics. The first step is time discretization.

Let $t_i, i = 0, \ldots, n$ such that $t_0 = 0, t_n = T$, denote regular mesh of time points with a mesh size $\Delta t = T/n$. As the $\Gamma(t)$ represents an uncountable set of points (S_t, t) laying on the optimal exercise boundary, then adopting time discretization turns the problem to solve (18) to looking for a finite set of points $\{t_i, S_i\}, i = 0, \ldots, n$, satisfying a discrete version of (18), where we just simplified notation of S_{t_i} to S_i by index shortening.

The time discretization being applied upon (18) yields the following set of fixed-point looking problems $S_i = \Psi(S_i; t_i, \Gamma(\eta)), i = 0, \ldots, n$, where Ψ expresses formally the right-hand side of (18) denoting its proper dependence upon variable S_i, but also upon parameters t_i and $\Gamma(\eta)$ thus making the problem rather involved. In order to overcome numerically a complication caused by function $\Gamma(\eta)$, $\eta \in (0, T)$, which is itself a solution of (18) being seeked, we propose the following iterative algorithm

1. Set initial approximation $\Gamma_0(\eta)$, $\eta \in (t_i, T)$, and set $k = 0$,
2. Increment current k by 1, and loop $t_i, i = 0, \ldots, n$,
   solve $S_{i,k} = \Psi(S_{i,k}; t_i, \Gamma_{k-1}(\eta))$, $k \geq 1$, until convergence of $S_{i,k}$ appears with given tolerance ϵ_1,
3. Build $\Gamma_k(\eta)$, $\eta \in (0, T)$ using available $\{t_i, S_{i,k}\}, i = 0, \ldots, n$ by spline approximation,
4. Check convergence, if $\|\Gamma_k - \Gamma_{k-1}\|_{L_2(0,T)} \leq \epsilon_2$ then *Stop* else *GoTo* step 2.

Our very first attempt to solve $S_i = \Psi(S_i; t_i, \Gamma(\eta))$ for a given t_i, which is a core of the step 2. of the given algorithm, was simply lured by computational power of sw Mathematica to solve it just by two commands

```
eq6318 = S[t] == eqSt := K + # Exp[-q (T - t)] CDF[
     NormalDistribution[0,
      1], -(-Log[#/K] + \[Beta]2 (T - t))/(\[Sigma] Sqrt[T - t])] -
   K Exp[-r (T - t)] CDF[
     NormalDistribution[0,
      1], -(-Log[#/K] + \[Beta]1 (T - t))/(\[Sigma] Sqrt[T - t])] -
   K r Integrate[
     Exp[-r (\[Eta] - t)] (1 -
        CDF[NormalDistribution[0,
          1], (Log[#/
              S\[Eta][\[Eta]]] + \[Beta]1 (\[Eta] -
               t))/(\[Sigma] Sqrt[\[Eta] - t])]), {\[Eta], t, T}] +
   +q # Integrate[
     Exp[-q (\[Eta] - t)] (1 -
        CDF[NormalDistribution[0,
          1], (Log[#/
              S\[Eta][\[Eta]]] + \[Beta]2 (\[Eta] -
               t))/(\[Sigma] Sqrt[\[Eta] - t])]), {\[Eta], t, T}] &

FindRoot[eqSt[s] == s, {s, s1}]
```

However, it failed with a message `SystemException["MemoryAllocationFailure"]` signalizing an extreme demand of internal Mathematica procedures upon memory allocation when executing such complex command. It was simply over the memory capacity of our Dell Latitude E6510 computer, with 4 GB RAM, processor Intel(R) i7 CPU Q 740 @ 1.73 GHz, 64-b MS-Win 10 Pro. Hence, we turned to break the given algorithm into less complex steps and control manually their performance.

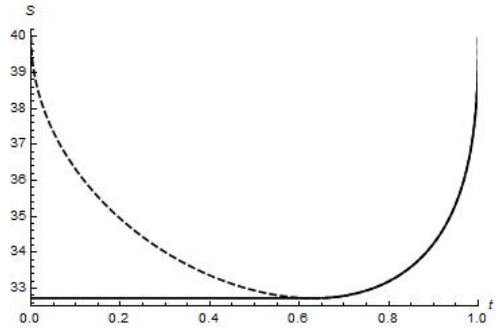
Following [4], Chapter 6.5, we also know some advanced information about optimal exercise boundary $\Gamma(t)$

$$\Gamma(T) = \min(\frac{rK}{q}, K), \tag{19}$$

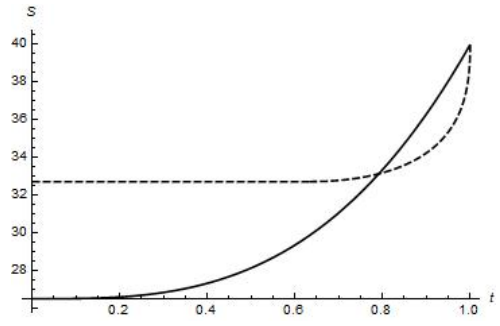
asymptotic expression of $\Gamma(t)$ near $t = T$ in case of $q = 0$

$$\frac{K - \Gamma(t)}{K} \approx \sigma\sqrt{(T-t)|\ln(T-t)|}, \tag{20}$$

and also, $\Gamma : t \to \Gamma(t)$ is a convex and monotonne increasing function on $(0, T)$.



(a) $\Gamma_0(t)$ – asymptotic-derived (full-line), $\Gamma(t)_{|asymp}$ – given by (20) (dashed).



(b) $\Gamma_0(t)$ – variants: 1) ad-hoc (full-line), 2) asymptotic-derived (dashed).

Figure 2: Initial variants of optimal exercise boundary $\Gamma_0(t), t \in (0, T)$.

For numerical determination of the optimal exercise boundary $\Gamma(t)$, we select the American put option with following data: $K = 40, r = 0.05, \sigma = 0.3, S \in (0, 65], T = 1, q = 0$. In Figure 2, we show two initial variants of $\Gamma_0(t)$ constructed either by using asymptotic expression (20), or by ad-hoc way using a cubic polynomial with proper interpolation conditions maintaining $\Gamma(T) = K$ by (19) and both qualitative properties, i.e. monotonicity and convexity of $\Gamma_0(t)$ on $(0, T)$. $\Gamma(t)_{|asymp}$ denotes a function calculated by expression (20) on whole interval $(0, T)$, which yields a convex function but violating monotonicity globally. It is depicted in Figure 2a by dotted line. To remedy this defect, we found a minimum point $t_{\min}$ of function (20) providing the minimum value $\Gamma_{\min}$, and make a prolongation from $t = 0$ to $t_{\min}$ by a horizontal line keeping the calculated minimal value $\Gamma_{\min}$, as it is depicted in Figure 2a, too, but in full-line.

Assumption $q = 0$, i.e. the case of non-dividend paying option, simplifies the equation (18) so that the first term is modified and the last one is dropped. Hence, we obtain another set of fixed-point looking problems $S_i = \psi(S_i; t_i, \Gamma(\eta)), i = 0, \ldots, n$, where ψ expresses formally the right-hand side of (21), now

$$\begin{aligned}\Gamma(t) = K + \Gamma(t)\Phi(-\frac{-\ln\frac{\Gamma(t)}{K} + \beta_2(T-t)}{\sigma\sqrt{T-t}}) - Ke^{-r(T-t)}\Phi(-\frac{-\ln\frac{\Gamma(t)}{K} + \beta_1(T-t)}{\sigma\sqrt{T-t}}) \\ - Kr\int_t^T e^{-r(\eta-t)}\big(1 - \Phi(\frac{\ln\frac{\Gamma(t)}{\Gamma(\eta)} + \beta_1(\eta-t)}{\sigma\sqrt{\eta-t}})\big)\mathrm{d}\eta\end{aligned} \tag{21}$$

where

$$\beta_1 = r - \frac{\sigma^2}{2}, \quad \beta_2 = r + \frac{\sigma^2}{2}.$$

Trapezoidal rule is used for approximation of integral $\int_t^T g(\eta)\mathrm{d}\eta$, which appears in (21), where $g(\eta)$ stands for an integrand taking the following form, keeping in mind that $t = t_i$ in particular case

$$g(\eta) = e^{-r(\eta-t)}\big(1 - \Phi(\frac{\ln\frac{\Gamma(t)}{\Gamma(\eta)} + \beta_1(\eta-t)}{\sigma\sqrt{\eta-t}})\big). \tag{22}$$

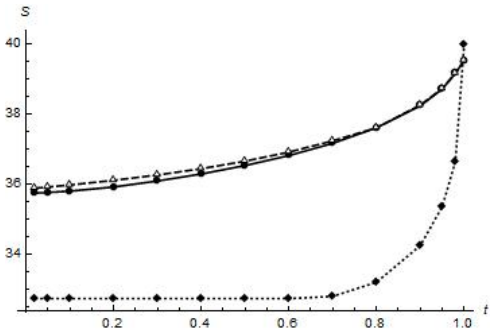
For trapezoidal rule, let m is a number of intervals covering (t_i, T) with $\Delta\eta = (T - t_i)/m$, thus providing

$$\int_{t_i}^T g(\eta)\mathrm{d}\eta \approx \frac{T-t_i}{2m}\Big(g(t_i) + 2\big\{\sum_{j=1}^{m-1} g(t_i + j\Delta\eta)\big\} + g(T)\Big) = G_i. \tag{23}$$

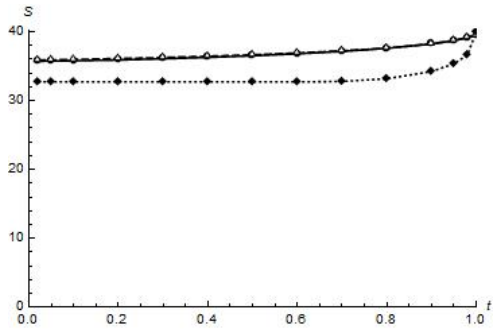
Now, we formulate the inner-most problem. For given $t_i, S_{i,0}$, find $S_{i,k}, k \geq 1$ such that (24) and (25) hold

$$\begin{aligned} S_{i,k} &= \psi(S_{i,k}; t_i, \Gamma_{k-1}(\eta)) \\ &= K + S_{i,k}\Phi(-\frac{-\ln\frac{S_{i,k}}{K} + \beta_2(T-t)}{\sigma\sqrt{T-t}}) - Ke^{-r(T-t)}\Phi(-\frac{-\ln\frac{S_{i,k}}{K} + \beta_1(T-t)}{\sigma\sqrt{T-t}}) - KrG_{i,k-1}, \end{aligned} \tag{24}$$

$$|S_{i,k} - \psi((S_{i,k}; t_i, \Gamma_{k-1}(\eta))| \leq \epsilon_1. \tag{25}$$



(a) Local comparison of computed optimal exercise boundary $\Gamma(t)$ (full-line and dashed), with asymptotic-derived $\Gamma_0(t)$ (dotted).



(b) Global comparison of computed $\Gamma(t)$ with asymptotic-derived $\Gamma_0(t)$.

Figure 3: Computed optimal exercise boundary $\Gamma(t), t \in (0, T)$.

Our results are depicted in Figure 3. Setting $\epsilon_1 = 10^{-3}$, we computed two variants of optimal exercise boundary starting from two initial functions $\Gamma_0(\eta)$ shown in Figure 2(b), in particular. Depicted points with their t-coordinates show the selected discrete set $\{t_i\}, i = 0, \ldots, n$, we have used for time discretization of the nonlinear Volterra integral equation (18). Making local comparison of both computed variants of optimal exercise boundary $\Gamma(t)$, as being given in Figure 3(a), we can conclude that different initial functions $\Gamma_0(t)$ cause almost negligible differences between computed optimal exercise boundary depicted by full-line and dashed line, respectively.

4 Conclusion

In the paper, we discuss briefly variational formulation of American option pricing problem, early exercise premium and early exercise boundary problem, which is a core of the paper. Numerical procedure proposed and coded in sw Mathematica seems to work good and represents a promising platform for our future research.

References

[1] Doffou, A.: Estimating the early exercise premium of American put index options, *Int. Journal of Banking and Finance* **6**(1) (2008), 31–47.

[2] Dueker, M.: Directly Measuring Early Exercise Premiums Using American and European S& P 500 Index Options,*Working Paper Series*. https://research.stlouisfed.org/wp/2002/2002-016.pdf.

[3] Fries, Ch.: *Mathematical finance – theory, modeling, implementation*. John Wiley & Sons Inc., Hoboken, New Jersey, 2007.

[4] Jiang, L.: *Mathematical Modeling and Methods of Option Pricing*. World Scientific Publ. Co., Singapore, 2005.

[5] Lukáš, L.: American Option Pricing Problem Formulated as Variational InequalityProblem. In: *Conf. Proceedings, 34-th Int. Conf. Mathematical Methods in Economics 2016*, (Kocourek, A., and Vavroušek, M., eds.). Technical Univ. of Liberec, Liberec, 2016, 512–517.

[6] Neftci, S.N.: *An Introduction to the Mathematics of Financial Derivatives*. 2-nd Ed., Academic Press, London, 2000.

[7] Pironneau, O., and Achdou, Y.: *Partial Differential Equations for Option Pricing.* In: *Mathematical Modeling and Numerical Methods in Finance*, Special Volume (Bensoussan, A., and Zhang, Q., eds) of *Handbook of Numerical Analysis.*, Vol. XV, (Ciarlet, P.G., ed.), North-Holland, Elsevier, Amsterdam, 2009.

[8] Stamicar, R., Ševčovič, D., and Chadam, J: The early exercise boundary for the American put near expiry: Numerical approximation, *Canadian Applied Mathematics Quarterly* **7**(4) (1999), 427–444.

Expected Coalition Influence under I-Fuzzy Setting: The Case of the Czech Parliament

Elena Mielcová[1]

Abstract.
The main aim of this article is to study coalition formation on real voting data in this case the data from the Chamber of Deputies of the Czech Parliament for two, 2006-2010 and 2010-2013 parliamentary periods. Power of political parties – players of this game – is measured by a Shapley-Shubik power index, Shapley-Shubik power index with a-priori union structure, and the blockability value. In order to cover uncertain character of real voting data, it is necessary to add uncertainty issues into the concept of coalitional values - namely the concept of Atanassov intuitionistic fuzzy (I-fuzzy) sets approach. All three ex-ante power indices for the period 2010-2013 are recalculated with respect to uncertainty values from the preceding period. Results of calculations are compared with overall voting outcome. Calculations show an improvement in results when uncertainty issues are considered.

Keywords: power indices with a-priori coalition structure, viability value, blockability value, parliamentary voting, I-fuzzy sets.

JEL classification: C71
AMS classification: 91A80

1 Introduction

In a voting game, the influence of players is commonly measured by power indices. However, a-priori coalition formation substantially changes influence of players. For the case of a-priori coalition structure, declared for example by coalition treaty, the correct way to evaluate the power is to compute so-called power indices with a-priori union structure. The idea of power index with a-priori coalition structure was introduced by Owen [8]. Kojima and Inohara [5][6] introduced different method for comparison of coalition influence of characteristic function games; they defined blockability and viability relations in order to compare influence of players as well as a blockability and a viability value evaluating a coalition influence by a real number.

The main aim of this text is to study a coalition formation influence in voting on outcome of voting and on selected indices - namely a Shapley-Shubik power index, a Shapley-Shubik power index with a-priori union structure, and a blockability value. Analysis is performed on real voting data in this case the data from the Chamber of Deputies of the Czech Parliament for two parliamentary periods. In order to cover uncertain character of real voting data, the uncertainty issues were studied by Aranassov intuitionistic fuzzy sets [1], also due to terminological difficulties called I-fuzzy sets [3]. Results of calculations are compared with overall voting outcome.

2 Preliminaries

The triple $G = [N, \mathbf{w}, q]$ is called a weighted voting game, where $N = \{1, 2, \dots n\}$ is a set of players, q is a quota and $\mathbf{w} = \{w_1, w_2, \dots w_n\}$ is a set of weights of respective players. A coalition $T \subset N$ is winning if total weight of players in the coalition meets or exceeds the quota q, $\sum_{i \in T} w_i \geq q$. The characteristic function of weighted voting game v can obtain only values 0 or 1; $v(T) = 1$ if T wins and $v(T) = 0$ when T loses. Moreover, the empty coalition never wins: $v(\emptyset) = 0$, the grand coalition always wins: $v(N) = 1$, and any superset of winning coalition also wins [4].

In order to evaluate the power of players in voting games, wide spectrum of different mathematical tools and techniques were introduced. The one, which is the standardly used and known, is the Shapley-Shubik power index [11]. This power index was designed such that it serve to a-priori evaluation of the power division among bodies in committee system under general settings: it is possible to use it not only in standard voting games, but also for example in multi-cameral systems or systems with restrictions. The derivation of the member's Shapley-Shubik power index is based on the number of cases when the member is in ordering pivotal: Let Π^p be a set of all $n!$ permutations of N. For every $\pi^p \in \Pi^p$ of the form $\pi^p = (\pi^p_1, \pi^p_2, \dots, \pi^p_n)$ there exists a unique k such that coalition $\{\pi^p_1, \pi^p_2, \dots, \pi^p_k\}$ wins, and $\{\pi^p_1, \pi^p_2, \dots, \pi^p_{(k-1)}\}$ does not win. Player π^p_k is the pivot of permutation π^p. Let Π^p_l denote the set of all permutations with pivot l. Then the Shapley-Shubik power index of player l can be

[1]Silesian University in Opava, School of Business Administration in Karviná, Univerzitní náměstí 1934/3, Karviná, mielcova@opf.slu.cz

expressed as [11]:

$$\phi_l = \frac{|\Pi_l^p|}{|\Pi^p|} = \frac{|\Pi_l^p|}{n!} \tag{1}$$

where the cardinality of sets Π_l^p, and Π^p is denoted by $|\Pi_l^p|$, and $|\Pi^p|$, respectively.

This approach is based on original Shapley idea of evaluation of game players by a value, now known as a Shapley value [10] simplified for voting games:

$$\phi_l = \sum_{\substack{l \in T \subset N \\ T \text{ winning} \\ T \backslash l \text{ losing}}} \frac{(t-1)!(n-t)!}{n!} \tag{2}$$

where summation is done through all winning coalitions $T \subset N$ containing player l such that a coalition that is created from T by omission of player l (denoted $T \backslash l$) is losing. The cardinality of the set T is denoted by t.

For the case of a-priori coalition structure, (for example in the case of coalition treaty), Owen [8][9] proposed to use so-called power index of the committee with a-priori union structure. For a voting game $[N, \mathbf{w}, q]$ let $\phi_i(q; w_1, w_2, \ldots, w_n)$ be the power index of i-th member of this voting game. Let's consider that members are creating m unions (coalitions) with weights $t_1, t_2, \ldots t_m$. Every union is composed of several members; each of them consists of at least one member. Denote $\mathbf{w}^j$ be the set of members in j-th union. Let's consider two-level processes: the first level process is an inter-union committee game $\{q; t_1, t_2, \ldots t_n\} = \{q; \mathbf{t}\}$ with the power of j-th coalition $\phi_j(q; \mathbf{t})$. The second level process covers the m internal union second level subgames $\{q; t_1, t_2, \ldots t_{j-1}, \mathbf{w}^j, t_{j+1}, \ldots t_m\} = \{q; \mathbf{t}; \mathbf{w}^j\}$ with the power index of i-th member of the j-th coalition $\phi_{ij}(q; \mathbf{t}; \mathbf{w}^j)$. Then the power index of the member i of the union j in the committee with a-priori union structure is

$$\phi_i(q; \mathbf{w}, \mathbf{t}) = \phi_j(q; \mathbf{t}) \frac{\phi_{ij}(q; \mathbf{t}; \mathbf{w}^j)}{\sum_{k \in K_j} \phi_{kj}(q; \mathbf{t}; \mathbf{w}^j)}, \tag{3}$$

where k is going through the set K_j containing all members in union j.

Another possibility how to compare coalition influence, was introduced by Kojima and Inohara [5] in a form of a blockability value, and a viability value. These two values assign a player with a real number; the bigger the number, the higher the respective coalition power. The blockability value of the coalition $S \subset N$ in the case of voting games can be expressed as:

$$B_S(N, v) = \frac{\sum_{T \subset N} v(T) - B^*(S)}{\sum_{T \subset N} v(T)} \tag{4}$$

where $B^*(S) = \sum_{T \subset N} v(T \backslash S)$. Similarly, the viability value of the coalition is expressed as:

$$V_S(N, v) = \frac{V^*(S)}{V^*(N)} \tag{5}$$

where $B^*(S) = \sum_{T \subset N} v(S \backslash T)$.

According to [2] [7] the term $(t-1)!(n-t)!/n!$ can be considered as probability of occurrence of a given coalition. Similarly, the concept of blockability and viability assumes uniform distribution of coalition formation probability. For illustration, Figure 1 contains comparison of coalition formation probability for five players for Shapley value as well as for blockability value.

As already mentioned, the concept of I-fuzzy coalitions was chosen to describe vagueness in Parliamentary voting, mainly because of ambivalence in members' decisions. Members' decisions are influencing party decisions and political parties are main players in the weighted voting game.

Let $N = \{1, 2 \ldots n\}$ be a set of n players. An I-fuzzy coalition $\tilde{C}$ is given by vectors $\tilde{C} = \langle \mu^C, \nu^C \rangle$ with coordinates $\mu^C = (\mu_1^C, \mu_2^C, \ldots \mu_n^C)$ and $\nu^C = (\nu_1^C, \nu_2^C, \ldots \nu_n^C)$ such that $0 \leq \mu_i^C + \nu_i^C \leq 1$ for all $i \in N$. A crisp coalition $S \subset N$ in I-fuzzy notation is represented by I-fuzzy coalition $\tilde{C}^S$ such that $\tilde{C}^S = \langle \mu^S, \nu^S \rangle$ for which $\mu_i^S = 1$ for all $i \in S$ and $\mu_i^S = 0$ for all $i \notin S$ while $\nu_i^S = 1$ for all $i \notin S$ and $\nu_i^S = 0$ for all $i \in S$.

Let T be a crisp coalition such that $T \subset N$. Let $\tilde{C}$ be an I-fuzzy coalition. Let e^T be a fuzzy coalition created from $\tilde{C}$ such membership and nonmembership function are at basic levels for players not in coalition T, other membership and nonmembership functions are unchanged. The Shapley value of the coalition $K \subset N$ for the referral coalition $\tilde{C}$ can be expressed as:

$$\Phi_i(v)_{\tilde{S}} = \sum_{i \in T \subset N} \frac{(t-1)!(n-t)!}{t!} \big(v\big(\sum_{j \in T} \tilde{S}_j e_j^T \big) - \big(v\big(\sum_{j \in T \backslash i} \tilde{S}_j e_j^T \big) \big) \big) \tag{6}$$

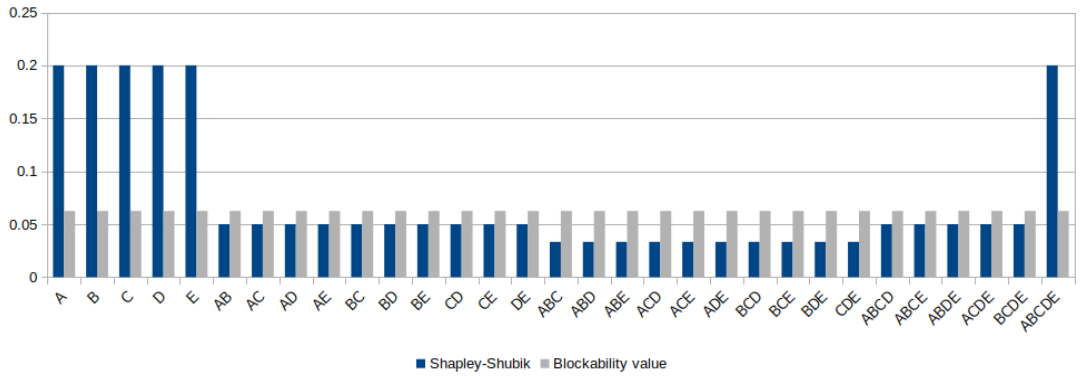


Figure 1 Depiction of expected probability of distributions for calculation of a Shapley-Shubik power index and a blockability value for five players.

The blockability value of the coalition $K \subset N$ for the referral coalition $\tilde{C}$ can be expressed as:

$$B_K = \frac{\sum_{T \subset N} v(e^T) - B^*(K)}{\sum_{T \subset N} v(e^T)} \tag{7}$$

where $B^*(K) = \sum_{T \subset N} v(e^T \backslash e^K)$. In real-world situation the referral coalition is considered to be a 'typical' coalition for a played game (for example in voting it can be the most probable coalition or an announced coalitional partnership).

In order to calculate a characteristic function in the case of the Chamber of Deputies of the Czech Parliament, a probabilistic approach was used. Values $v(e^T)$ and $v(e^T \backslash e^K)$ are calculated such that: if the sum of participation levels of party members in coalition T (for simplicity called $\min(T)$) is greater than 100, then $v(T) = 1$; if the sum of participation levels of party members in coalition T plus all values of uncertainty levels of all political parties in T (for simplicity called $\max(T)$) is smaller than 101, then $v(T) = 0$; otherwise the respective value $v(T)$ is calculating as the value of $v(T) = 1 - F(100)$, where $F(100)$ is the value of a cumulative distribution function for the continuous uniform distribution.

3 Data Description

This analysis is based on the data from the Chamber of Deputies of the Czech Parliament for two parliamentary periods, namely 2006-2010, and 2010-2013 periods. Information of votes are available at the official web site of the Chamber of Deputies of the Czech Parliament URL: www.psp.cz. Data for the 2006-2010 parliamentary period are composed from 8740 voting outcomes for each legislator; data for the 2010-2013 parliamentary period are composed from 5898 voting outcomes. The outcome of every vote for each member is one possibility from the set {yes, no, abstain, absent}.

During the first 2006-2010 period there were five political parties operating in the Chamber of Deputies of the Czech Parliament: Civic Democratic Party (ODS), Christian and Democratic Union - Czechoslovak People's Party (KDU-CSL), Green Party (SZ), Czech Social Democratic Party (CSSD), Communist Party of Bohemia and Moravia (KSCM). Three political parties ODS, KDU-CSL, and SZ created governmental coalition; other two political parties CSSD and KSCM stayed in opposition. Distribution of seats in the 2006-2010 Chamber of Deputies of the Czech Republic together with party voting success index is given in Table 1.

Political Party	ODS	CSSD	KSCM	KDU-CSL	SZ
Seats	81	74	26	13	6
Party Success	0.808	0.812	0.721	0.749	0.678

Table 1 Distribution of seats and Party Success index in the 2006-2010 Chamber of Deputies of the Czech Republic

During the second, 2010-2013 parliamentary period there were five political parties operating in the Chamber of Deputies of the Czech Parliament: Civic Democratic Party (ODS), TOP09, Veci verejne (VV), Czech Social Democratic Party (CSSD), and Communist Party of Bohemia and Moravia (KSCM). Three political parties – ODS, TOP09, and VV – created governmental coalition; other two political parties CSSD and KSCM stayed in opposition. Distribution of seats in the 2010-2013 Chamber of Deputies of the Czech Republic together with party voting success index is given in Table 2.

Political Party	CSSD	ODS	TOP09	KSCM	VV
Seats	56	53	41	26	24
Party Success	0.592	0.943	0.952	0.567	0.831

Table 2 Distribution of seats and Party Success index in the 2010-2013 Chamber of Deputies of the Czech Republic

4 Results and Discussion

The standard way how to measure future power of players – in our case political parties – is to evaluate them by predefined indices. Table 3 gives an a-priori evaluation of political parties in 2006-2010 Chamber of deputies using Shapley value, Shapley value with a-priori coalition structure (for expected governmental coalition ODS, KDU-CSL, and SZ) and blockability value. Viability value for the governmental coalition was 0. All correlation coefficients of respective indices with Party Success index are positive and quite high. However, they are not statistically significant at 5% level of significance.

Political party	ODS	CSSD	KSCM	KDU-CSL	SZ	Correlation
Shapley-Shubik power index	0.367	0.283	0.283	0.033	0.033	0.69
Shapley-Shubik with a-priori union structure	0.121	0.333	0.333	0.112	0.101	0.25
Blockability value	0.600	0.467	0.467	0.067	0.067	0.69

Table 3 Results of a-priori values for 2006-2010 Chamber of Deputies of the Czech Republic. Last column gives a correlation coefficient of the respective index with values of Party Success.

Analogically, respective results for political parties present in 2010-2013 Chamber of Deputies are given in Table 4. In the case of Shapley value with a-priori coalition structure the expected governmental coalition was composed of ODS, TOP09, and VV; viability value of this coalition was 0.1875. The correlation coefficient between a-priori Shapley-Shubik power index and a Party Success is the same as for Blocking value, it reaches value of -0.046 and this correlation coefficient is not statistically significant (p=0.941). Similarly, the correlation coefficient between Shapley-Shubik power index with a-priori union structure and a Party Success value is equal to 0.868 (p=0.056).

Political party	CSSD	ODS	TOP09	KSCM	VV	Correlation
Shapley-Shubik power index	0.300	0.300	0.133	0.133	0.133	-0.046
Shapley-Shubik with a-priori union structure	0	0.530	0.235	0	0.235	0.868
Blockability value	0.5	0.5	0.25	0.25	0.25	-0.046

Table 4 Results of a-priori values for 2010-2013 Chamber of Deputies of the Czech Republic. Last column gives a correlation coefficient of the respective index with values of Party Success.

One of the aspects that could influence of future success is the probability of coalition creation – usually coalitions do not have expected probabilities. For illustration, Figures 2 and 3 give relative frequencies of voting in coalition with respect to overall voting.

Relative frequencies in Figures 2, and 3 can be considered as real probability distributions over all possible coalitions. These probability distributions are symmetric (e.g. coalition ABC implies creation of coalition DE, there is no empty coalition depicted in figures), and for both periods reveal that the most probable coalition is the grand coalition followed by governmental coalition. In the next calculations, the governmental coalition was set to be a referral coalition. Membership values and nonmembership values of political parties in governmental coalition were calculated with respect to the relative frequency, results are given in Table 5.

Calculated values of membership and nonmembership values for 2006-2010 Chamber of Deputies of the Czech Parliament were used to predict power distribution in 2010-2013 Chamber of Deputies. As not all political parties were present in Parliament during both periods, membership and nonmembership values of such political parties were set as average values of all political parties in precedent period. Results of calculated IF indices together are given in Table 6. For all calculated a-priori indices the correlations between index and overall voting outcome increased, even though only one of them is statistically significant at 5% level of significance, namely Shapley-Shubik power index with a-priori union structure based on a-priori governmental coalition of ODS, TOP09, and VV.

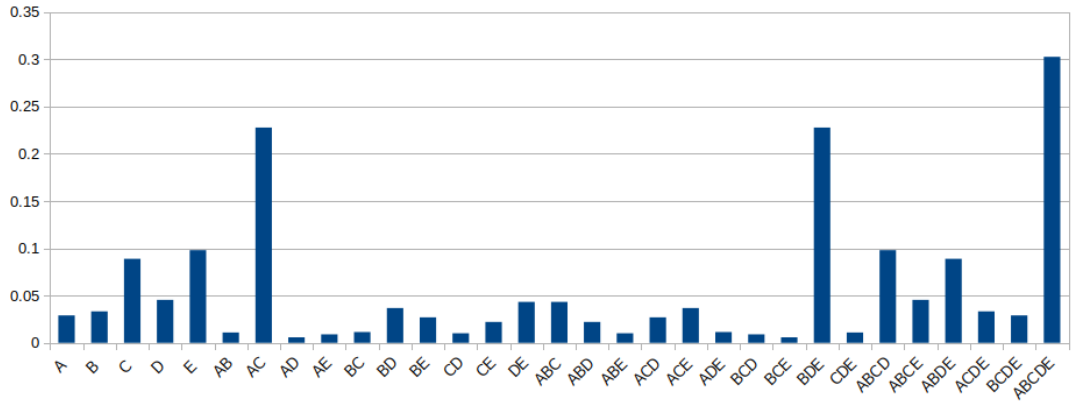


Figure 2 Probability of coalition creation for the 2006-2010 Chamber of Deputies of the Czech Parliament; probabilities are calculated as a relative frequencies. A = ODS, B = CSSD, C = KSCM, D = KDU-CSL, E = SZ.

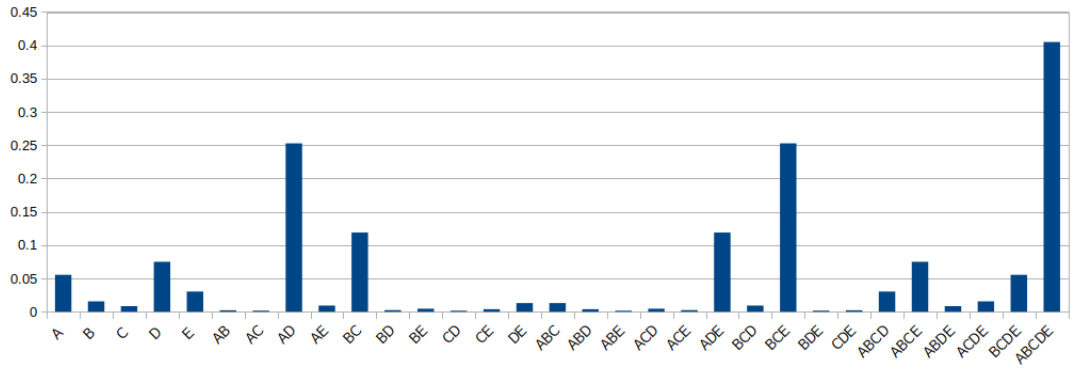


Figure 3 Probability of coalition creation for the 2010-2013 Chamber of Deputies of the Czech Parliament; probabilities are calculated as a relative frequencies. A = CSSD, B = ODS, C = TOP09, D = KSCM, E = VV.

Political Party	ODS	CSSD	KSCM	KDU-CSL	SZ
Membership Value	0.8096	0.7743	0.8668	0.7197	0.6229
Nonmembership Value	0.0288	0.0434	0.0332	0.0521	0.0340

Table 5 Calculated membership and nonmembership values for 2006-2010 Chamber of Deputies of the Czech Parliament

Political party	CSSD	ODS	TOP09	KSCM	VV	Correlation
Membership Value	0.7743	0.8096	0.7587	0.8668	0.7587	
Nonmembership Value	0.0434	0.0288	0.0383	0.0332	0.0383	
Shapley-Shubik power index	0.2923	0.2814	0.2181	0.1183	0.0899	0.1564
Shapley-Shubik with a-priori union structure	0.0672	0.4133	0.3200	0.0672	0.1320	0.8918
Blocking value	0.4425	0.5863	0.3777	0.1797	0.1231	0.3546

Table 6 Results of calculations of a-priori values for 2010-2013 Chamber of Deputies of the Czech Republic with incorporated uncertainty issues. Last column gives a correlation coefficient of the respective index with values of Party Success.

5 Conclusion

The main aim of this text was to show that a coalition formation is influencing voting outcome and that this influence can be partially incorporated into selected a-priori power indices such that they could better reflect future outcome of voting. Analysis was performed on real voting data in this case the data from the Chamber of Deputies of the Czech Parliament for two parliamentary periods. Voting outcome was measured by success index, power of players was evaluated by three indices, namely Shapley-Shubik power index, blockability value, Shapley-Shubik power index with a-priori union structure based on formed governmental coalition. For both observed period (2006-2010, and 2010-2013 parliamentary periods) the correlation coefficient between indices and success index is quite small with the exception of Shapley-Shubik power index with a-priori union structure in the second period (which is still not statistically significant at 5% level of significance).

In order to cover uncertain character of real voting data, the uncertainty issues represented by I-fuzzy sets were incorporated into calculated a-priori values. The first period served as a reference period for calculation of membership and nonmembership values. These values, together with seats distribution were taken as input values for calculation of a-priori indices for the second period. For all three calculated indices the correlation coefficient increased, the Shapley-Shubik power index with a-priori union structure is statistically significant at 5% level of significance, other two indices are still not statistically significant.

Acknowledgements

This paper was supported by the Ministry of Education, Youth and Sports Czech Republic within the Institutional Support for Long-Term Development of a Research Organization in 2017.

References

[1] Atanassov, K. T.: Intuitionistic Fuzzy Sets, *Fuzzy Sets and Systems*, **20**(1) (1986), 87–96.

[2] Bilal S., Albuquerque P., Hosli M. O.: *The Probability of Coalition Formation: Spatial Voting Power Indices, The ECPR General Conference/Pan-European IR Conference*, Canterbury, England, 6-8 September 2001, Madison, Wisconsin, AEI http://aei.pitt.edu/2052/1/001591_1.pdf. Accessed May 1, 2017.

[3] Dubois, D., Gottwald, S., Hajek, P., Kacprzyk, J., Prade, H.: Terminological Difficulties in Fuzzy Set Theory – The Case of "Intuitionistic Fuzzy Sets", *Fuzzy Sets and Systems*, **156**(3) (2005), 485–491.

[4] Hu, X.: An asymmetric ShapleyShubik power index, *International Journal of Game Theory*, **34** (2006), 229-240.

[5] Kojima, K, Inohara, T.: Methods for comparison of coalition influence on games in characteristic function form, *Applied Mathematics and Computation*, **217** (2010), 4047–4050.

[6] K. Kojima, K, Inohara, T.: Coalition values derived from methods for comparison of coalition influence for games in characteristic function form, *Applied Mathematics and Computation*, **219** (2012), 1345–1353.

[7] Owen, G.: Multilinear Extensions of Games, *Management Science*, **18**(5) (1972), 64–79.

[8] Owen, G.: Values of Games with a priori Unions. In: Henn, R., Moeschlin, O. (eds.): *Essays in mathematical economics and game theory.* Fields Institute, Berlin, 1977, 76–88.

[9] Owen, G.: *Game theory*. 3rd ed. San Diego: Academic Press, 1995.

[10] Shapley, L. S.: A value for n-Person Games. In: Kuhn, H. W., and Tucker A. W. (eds.): *Contribution to the Theory of Games* vol. 2 (1953), Princeton University Press, Princeton, New Jersey, 307–317.

[11] Shapley, L. S., Shubik, M.: A Method for Evaluating the Distribution of Power in a Committee System, *American Political Science Review*, **48** (1954), 784–792.

Comparison of rankings of decision alternatives based on the omega function and the prospect theory

Ewa Michalska[1], Renata Dudzińska-Baryła[2]

Abstract. Comparing decision alternatives in the prospect theory the deviations of outcomes from the reference point are evaluated. The similar approach is applied in the idea of the omega function, which is a ratio of the expected value of the relative gains and expected value of the relative losses, calculated in relation to the threshold level (reference point). Moreover, these criteria are defined for continuous distributions, which allows to include full information about the distribution of outcomes. In this paper the random rates of return of stocks quoted on Warsaw Stock Exchange are the decision alternatives. The aim of this paper is to examine how the selection of the probability distribution of the random rates of return, and the way the rates of return are defined, influence the rankings of stocks created on the basis of the omega ratio and the prospect theory. Various distributions as well as various ways of defining the rate of return are considered.

Keywords: prospect theory, omega ratio, omega function, performance measure, continuous probability distribution.

JEL Classification: G11
AMS Classification: 91B06, 91G99

1 Introduction

Fundamental concept of the prospect theory bases on the behavioral evaluation of a random decision alternative which depends on information how individuals evaluate potential gains and losses with respect to the reference point [3]. A similar approach is used in evaluating a random decision alternative using the omega function, which is the quotient of the expected relative gains and the expected relative losses measured against the threshold (reference point). The value of the omega function for a fixed reference point is a performance measure of an alternative. The omega function was defined by Shadwick and Keating [9] both for discrete and continuous distributions, but in prospect theory (PT) a formula for evaluating the decision alternative described by continuous distribution was proposed first time in 2008 [8]. In both measures (PT and omega) full information on the distribution can be considered. The purpose of this paper is to analyze the consistency of the rankings of decision alternatives with given continuous distributions and depending on the given way of defining them. In empirical studies the random rates of return of selected shares listed on the Warsaw Stock Exchange were used as the decision alternatives.

2 Omega function for continuous distribution

Searching for the performance measure taking into account all the information about the distribution of the random variable led to proposition of omega function [9]. If $F(x)$ denotes the cumulative distribution of the random variable X the omega function is defined as follows:

$$\Omega_L = \frac{\int_L^{+\infty} (1 - F(x))dx}{\int_{-\infty}^{L} F(x)dx} \tag{1}$$

where L is a variable threshold level. If a variable X stands for a random rate of return and the threshold L is a level of the rate of return determined by the decision-maker then the value of the omega function is the quotient of the expected value of gains and expected value of losses (relative to the threshold L).

[1] University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, email: ewa.michalska@ue.katowice.pl
[2] University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, email: dudzinska@ue.katowice.pl

The analytic form of the omega function depends on the assumed distribution of the random variable and its determination is often very difficult. A good starting point for determining both the analytical form of the omega function and the value of the omega function for continuous distributions with the given density function $f(x)$ of the random variable X is the following relation:

$$\Omega_L = \frac{\int_L^{+\infty} (x-L) f(x)\, dx}{\int_{-\infty}^{L} (L-x) f(x)\, dx} \tag{2}$$

If $f(x)$ is the density function of the normal distribution $\mathrm{N}(\mu,\sigma)$ than the omega function (2) has the following form [5]:

$$\Omega_L = \frac{(\mu - L)\,\Phi\left(-\frac{L-\mu}{\sigma}\right) + \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{L-\mu}{\sigma}\right)^2\right)}{(L-\mu)\,\Phi\left(\frac{L-\mu}{\sigma}\right) + \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{L-\mu}{\sigma}\right)^2\right)}, \tag{3}$$

where Φ denotes the cumulative distribution function of the standard normal distribution, μ the expected value and σ the standard deviation of random variable X. In the paper [6] the authors also propose an analytic form of the omega function for a random variable with a uniform distribution and a random variable being a linear transformation of a random variable with a uniform distribution.

Another type of continuous distribution often used for random rates of return is the normal inverse Gaussian distribution (NIG) belonging to the class of generalized hyperbolic distributions. Piasecki and Tomasik [7] failed to reject the hypothesis of the NIG distribution for 93% of the analyzed rates of return. The density function of the random variable with the $\mathrm{NIG}(\alpha,\beta,\delta,\tau)$ distribution is:

$$f_{\mathrm{NIG}}(x) = \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2 + (x-\tau)^2}\right)}{\pi\sqrt{\delta^2 + (x-\tau)^2}} \exp\left[\left(\delta\sqrt{\alpha^2 - \beta^2}\right) + \beta(x-\tau)\right], \tag{4}$$

where: α is a tail heaviness, β is an asymmetry parameter, δ is a scale parameter, τ is the location parameter, and $K_1(\cdot)$ denotes a modified Bessel function of the third kind. Determining the analytical form of the omega function for the NIG distribution is not a simple task given the complicated formula of the density function. In this situation it is better to use the integral formula (2) especially if calculations of the value of omega functions for fixed thresholds are supported by mathematical software (e.g. R or Mathematica).

Following the principle of preference based on the omega ratio we say that the random decision alternative D1 is preferred over D2 (written as $\mathrm{D1} \succ_{\Omega_L} \mathrm{D2}$) for a fixed threshold L if and only if an inequality $\Omega_L(\mathrm{D1}) > \Omega_L(\mathrm{D2})$ holds.

3 Valuation of decision alternatives with continuous distribution based on the prospect theory

The authors of the prospect theory Kahneman and Tversky [3] have extended the concept of the expected utility with psychological aspects of decision making. The absolute outcomes of the decision x_i (final assets or total wealth) were replaced with $z_i = x_i - L$, gains and losses relative to a certain reference point L. In addition, the utility function has been replaced by a value function $v(z)$ that reflects the risk-aversion behavior for gains and the risk-seeking behavior for losses (concave in the domain of gains and convex in the domain of losses). This function is steeper for losses than for gains, which expresses the fact that loss hurts more than gain of the same value satisfies. The prospect theory also relates decision weights to stated probabilities by the weighting probability function $w(p)$, where p is a probability of a relative outcome. This function models the behavior of decision makers who tend to overestimate very small probabilities and underestimate large ones. The most frequently cited forms of these functions are [10]:

- the value function $v(z)=\begin{cases} z^{0,88} & z \geq 0 \\ -2,25(-z)^{0,88} & z<0 \end{cases}$ (5)

- the probability weighting function $w(p)=\dfrac{p^{\gamma}}{\left(p^{\gamma}+(1-p)^{\gamma}\right)^{1/\gamma}}$ (6)

and the value of γ in the study of Tversky and Kahneman [10] was 0.69 for losses and 0.61 for gains.

Limitation of the original prospect theory [3] e.g. the possibility of evaluating random decision alternative with at most two non-zero outcomes, has been overcome. Rieger and Wang [8] proposed a formula for evaluating alternative with multiple outcomes can be written as follows

$$\mathrm{PT}_L=\frac{\sum_{i=1}^{n} w(p_i)\cdot v(z_i)}{\sum_{i=1}^{n} w(p_i)} \tag{7}$$

The weight normalization[3] in (7) is intended to prevent the divergence to $+\infty$. Generalizing the formula (7) for the decision alternative given by the continuous distribution with density function $f(z)$, Rieger and Wang [8] proposed the following analytical form:

$$\mathrm{PT}_L=\frac{\int_{-\infty}^{\infty} v(z)\cdot\left(f(z)\right)^{\gamma} dz}{\int_{-\infty}^{\infty}\left(f(z)\right)^{\gamma} dz} \tag{8}$$

In particular, if the value function is given by (5), PT value is calculated as follows [1, 2]

$$\begin{aligned}\mathrm{PT}_L &= \frac{\int_{-\infty}^{0} -2,25(-z)^{0,88}\left(f(z)\right)^{0,69} dz+\int_{0}^{+\infty} z^{0,88}\left(f(z)\right)^{0,61} dz}{\int_{-\infty}^{0}\left(f(z)\right)^{0,69} dz+\int_{0}^{+\infty}\left(f(z)\right)^{0,61} dz}= \\ &= \frac{\int_{-\infty}^{0} -2,25(-z)^{0,88}\left(f(z)\right)^{0,69} dz}{\int_{-\infty}^{0}\left(f(z)\right)^{0,69} dz+\int_{0}^{+\infty}\left(f(z)\right)^{0,61} dz}+\frac{\int_{0}^{+\infty} z^{0,88}\left(f(z)\right)^{0,61} dz}{\int_{-\infty}^{0}\left(f(z)\right)^{0,69} dz+\int_{0}^{+\infty}\left(f(z)\right)^{0,61} dz}= \\ &= \mathrm{PT}_L^{-}+\mathrm{PT}_L^{+}\end{aligned} \tag{9}$$

The PT value can be expressed as a sum of the valuation of losses PT_L^{-} and the valuation of gains PT_L^{+}. Following the principle of preference based on the PT value we say that the random decision alternative D1 is preferred over D2 (written as $\mathrm{D1}\succ_{\mathrm{PT}_L}\mathrm{D2}$) for a fixed reference point L if and only if an inequality $\mathrm{PT}_L(\mathrm{D1})>\mathrm{PT}_L(\mathrm{D2})$ holds.

4 Empirical analysis of the consistency of rankings

The purpose of the paper is to investigate whether the distribution of probabilities of random rates of return and the definition of the rate of return determine the companies' rankings based on the omega ratio and PT value. The research uses quotations of the 195 companies listed on Warsaw Stock Exchange in the first half of 2016 year (without any suspension in quotations). Based on these data, daily rates of return (R1) and logarithmic daily rates of return (R2) were calculated as follows:

[3] Karmarkar [4] introduced the weight normalization in the subjectively weighted utility.

$$\mathrm{R1}_t = \frac{C_t - C_{t-1}}{C_{t-1}}, \tag{10}$$

$$\mathrm{R2}_t = \ln\frac{C_t}{C_{t-1}}, \tag{11}$$

where C_t denotes share's quotation in moment t.

For random rates of return, two types of continuous distributions were considered - normal distribution (NORM) and normal inverse Gaussian distribution (NIG) with density function (4). For each stock and both R1 and R2 we estimated the parameters of the normal distribution (sample mean and standard deviation) and we tested the null hypothesis that empirical distribution is the same as estimated normal distribution (Anderson-Darling test with 0.01 significance level in the *nortest* Package of R software). The parameters of NIG distribution for rates of return R1 and R2 were estimated using *GeneralizedHyperbolic* Package. The Kolmogorov-Smirnov test with 0.01 significance level was used to test the null hypothesis that the empirical distribution is the same as the estimated NIG distribution. From all 195 stocks, 45 were selected for which we fail to reject the null hypothesis about the normal distribution as well as NIG distribution for both R1 and R2 rates of return. In following part rankings of these 45 stocks were built.

The evaluation of a random decision alternative based on the omega ratio and PT value requires assuming a threshold level L (reference point). In the research we use values L that differ by 0.01 and are within the interval $\langle -0.03, 0.02\rangle$ which is an intersection of ranges of observed rates of returns. For the rates of return R1 and R2, different threshold values and the assumed distribution (NORM or NIG), omega ratios and PT values were calculated. Subsequently, rankings of companies were created. We use Spearman's correlation coefficient as a measure of the interdependencies between the positions of shares in the rankings. The significance of the obtained coefficients at significance level 0.01 was also examined.

In order to investigate how the definition of the random rates of return influences the ranking, the rankings created for the assumed probability distribution (NORM or NIG) and the fixed threshold value (12 pairs of rankings based on the omega ratio and 12 pairs of rankings based on PT value) were compared. Table 1 shows the values of Spearman's rank correlation coefficients, which express the strength of the statistical dependence between the rankings created for different rates of return (R1 or R2), the assumed distribution (NORM or NIG) and using the omega ratios and PT values.

L	Correlation between $\Omega_L^{\mathrm{NORM}}(\mathrm{R1})$ and $\Omega_L^{\mathrm{NORM}}(\mathrm{R2})$	Correlation between $\Omega_L^{\mathrm{NIG}}(\mathrm{R1})$ and $\Omega_L^{\mathrm{NIG}}(\mathrm{R2})$	Correlation between $\mathrm{PT}_L^{\mathrm{NORM}}(\mathrm{R1})$ and $\mathrm{PT}_L^{\mathrm{NORM}}(\mathrm{R2})$	Correlation between $\mathrm{PT}_L^{\mathrm{NIG}}(\mathrm{R1})$ and $\mathrm{PT}_L^{\mathrm{NIG}}(\mathrm{R2})$
-0.03	0.9982	0.9959	0.9976	0.9788
-0.02	0.9996	0.9987	0.9980	0.9810
-0.01	0.9992	0.9993	0.9970	0.9854
0	0.9988	0.9989	0.9968	0.9879
0.01	0.9996	0.9997	0.9963	0.9822
0.02	0.9991	0.9978	0.9964	0.9798

Table 1 Spearman's rank correlation coefficients for rankings based on evaluation of rates of return for various way of defining rates of return (R1 or R2)

For fixed values of reference point, rankings based on the omega ratio do not depend on the definition of the rate of return (Spearman correlation coefficients ≈ 1), both for the NORM distribution as well as for the NIG distribution. We can observe a similar strength of dependence for rankings based on the PT values.

In order to investigate whether the assumed distribution (NORM or NIG) of the random rates of return influences the ranking we compared the rankings created for the assumed definition of rate of return (R1 or R2) and the fixed threshold value L (12 pairs of rankings based on the omega ratio and 12 pairs of rankings based on PT value). Table 2 shows the values of Spearman's rank correlation coefficients, which express the strength of the statistical dependence between the rankings created for different assumed distributions (NORM or NIG), using different definitions of rates of return (R1 or R2) and the omega ratio and PT values.

L	Correlation between $\Omega_L^{\mathrm{NORM}}(\mathrm{R1})$ and $\Omega_L^{\mathrm{NIG}}(\mathrm{R1})$	Correlation between $\Omega_L^{\mathrm{NORM}}(\mathrm{R2})$ and $\Omega_L^{\mathrm{NIG}}(\mathrm{R2})$	Correlation between $\mathrm{PT}_L^{\mathrm{NORM}}(\mathrm{R1})$ and $\mathrm{PT}_L^{\mathrm{NIG}}(\mathrm{R1})$	Correlation between $\mathrm{PT}_L^{\mathrm{NORM}}(\mathrm{R2})$ and $\mathrm{PT}_L^{\mathrm{NIG}}(\mathrm{R2})$
-0.03	0.9398	0.9419	0.6101	0.7047
-0.02	0.9750	0.9789	0.6827	0.7472
-0.01	0.9872	0.9884	0.7466	0.7851
0	0.9997	0.9999	0.8045	0.8292
0.01	0.9901	0.9893	0.8231	0.8460
0.02	0.9640	0.9589	0.8246	0.8672

Table 2 Spearman's rank correlation coefficients for rankings based on evaluation of rates of return for various distributions (NORM or NIG)

For all considered reference points, the correlation coefficients for rankings based on the omega ratio are higher than 0.93, which means that the making an assumption of a type of distribution (NORM or NIG) does not significantly affect the ranking of the shares. For rankings based on the PT values the correlation coefficients are lower than 0.87, but higher than 0.61, which means that the distribution type influences the ranking to a greater extent than when using the omega ratio.

In the next part, we investigate the interdependence between the rankings based on the omega ratios and the PT values for the assumed distribution and the definition of rate of return.

L	Correlation between $\Omega_L^{\mathrm{NORM}}(\mathrm{R1})$ and $\mathrm{PT}_L^{\mathrm{NORM}}(\mathrm{R1})$	Correlation between $\Omega_L^{\mathrm{NORM}}(\mathrm{R2})$ and $\mathrm{PT}_L^{\mathrm{NORM}}(\mathrm{R2})$	Correlation between $\Omega_L^{\mathrm{NIG}}(\mathrm{R1})$ and $\mathrm{PT}_L^{\mathrm{NIG}}(\mathrm{R1})$	Correlation between $\Omega_L^{\mathrm{NIG}}(\mathrm{R2})$ and $\mathrm{PT}_L^{\mathrm{NIG}}(\mathrm{R2})$
-0.03	0.8986	0.9032	0.5962	0.6909
-0.02	0.9441	0.9502	0.6667	0.7307
-0.01	0.9943	0.9941	0.7958	0.8341
0	0.5717	0.5565	0.6978	0.6509
0.01	-0.0378*	-0.0988*	0.2461*	0.1638*
0.02	-0.0315*	-0.0814*	0.3170*	0.2240*

* not significantly different from zero at the 1% level

Table 3 Spearman's rank correlation coefficients for rankings based on evaluation of rates of return using omega ratio and PT

For the NORM distribution of random rates of return for both R1 and R2, Spearman's correlation coefficients are relatively high for negative reference points L. On the other hand, for positive reference points, Spearman's rank correlation coefficients are not significantly different from zero, which indicates a lack of interdependence between these rankings. For the NIG distribution and negative values of reference point L, the Spearman's rank correlation coefficients are between 0.59 and 0.84. For positive values of reference point, rankings based on the omega ratio and PT value are not interdependent. The lack of interdependency between rankings for positive threshold levels can be explained by different forms of the omega ratio (quotient) and PT value (sum). If threshold levels are positive than most of outcomes are understood as losses. In such case the value of the denominator of the omega ratio increases, which can lead to larger changes in rankings relative to rankings based on PT value, than for negative threshold levels. However, confirmation of the reason for such situation requires further investigation.

5 Conclusions

The use of decision support tools often requires that certain assumptions are met. According to Markowitz theory, in the field of investment it is assumed that random variables are normally distributed. In this study we attempted to determine what are the consequences of assuming that random rates of return are normally

distributed when omega and PT are used to evaluate the decision alternatives (shares). Rankings of shares were compared with the rankings build on the bases of the omega and PT but assuming the NIG distribution (it was found that the majority of the random rates of return can be modeled by the NIG distribution [7]).

Comparing decision alternatives (random rates of return) valuated on the basis of the prospect theory, the choice of probability distribution influences the ranking, but it does not matter if we take the daily rates of return or the logarithmic daily rates of return. On the other hand, if a decision-maker evaluate the shares on the basis of the omega ratio, he or she do not need to think very carefully about the way the rate of return are defined or which distribution is assumed (NORM or NIG). It is also very important if the decision-maker chooses the objective (the omega ratio) or the subjective (the prospect theory) approach while evaluating the decision alternatives, especially when most relative outcomes are perceived as losses.

References

[1] Dudzińska-Baryła, D.: Ocena wariantów decyzyjnych o rozkładach ciągłych na gruncie teorii perspektywy. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach* **208** (2014), 7–18.

[2] Dudzińska-Baryła, D.: Stock Valuation Based on Prospect Theory for Continuous Distribution. In: *Conference Proceedings, 33rd International Conference Mathematical Methods in Economics* (Martinčik D., Ircingová J., Janeček P., eds.). University of West Bohemia, Plzeň, 2015, 150–155.

[3] Kahneman, D., Tversky, A.,: Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* **47** (1979), 263–291.

[4] Karmarkar, U.S.: Subjectively Weighted Utility: A Descriptive Extension of the Expected Utility Model. *Organizational Behavior and Human Performance* **21** (1978), 61–72.

[5] Michalska, E., Dudzińska-Baryła, R.: Wskaźnik omega w ocenie wariantów decyzyjnych o rozkładach ciągłych na przykładzie akcji notowanych na GPW w Warszawie. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach* **241** (2015), 112–124.

[6] Michalska, E., Kopańska-Bródka, D.: The Omega Function for Continuous Distribution, In: *Conference Proceedings, 33rd International Conference Mathematical Methods in Economics* (Martinčik D., Ircingová J., Janeček P., eds.). University of West Bohemia, Plzeň, 2015, 543–548.

[7] Piasecki, K., Tomasik, E.: *Rozkład stóp zwrotu z instrumentów polskiego rynku kapitałowego*. edu-Libri, Kraków-Warszawa, 2013.

[8] Rieger, M.O., Wang, M.: Prospect Theory for Continuous Distributions. *Journal of Risk and Uncertainty* **36** (2008), 83–102.

[9] Shadwick, W., Keating, C.: A Universal Performance Measure. *Journal of Performance Measurement* **6(3)** (2002), 59–84.

[10] Tversky, A., Kahneman, D.: Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* **5** (1992), 297–323.

Interpretation of Human Resource Management Data using Multiple Correspondence Analysis

Ondřej Mikulec[1]

Abstract. Ability to properly work with quantitative data is crucial for each segment of a company including Human Resource Management (HRM). People analytics represents a data-driven approach to HRM and implementation of quantitative models in decision processing of company´s workforce planning. In this work we present a combined way of using a multiple correspondence analysis as multivariate analysis method that integrates ideas from multidimensional scaling in the field of people analytics.
Several managers formulated ideas concerning possible dependencies and associations among length and frequency of absenteeism events, age of an employee, sex and employee´s designation (blue collar, white collar). These ideas are formulated as hypotheses and tested on data set summarizing information about more than five thousand employees of six production companies from Moravian-Silesian region during period 01/2015 - 12/2016. Associations and particular patterns among specified categorical variables are interpreted using Multiple Correspondence Analysis (MCA). The main objective of this study is to discuss obtained information with focus on implementation of the results in the decision processing of HRM. We also conclude that MCA represents appropriate statistical tool for visualizing HRM data.

Keywords: Multiple correspondence analysis, human resource management, people analytics, absenteeism.

JEL Classification: C3, M50
AMS Classification: 91

1 Introduction

This study presents new empirical evidence on the associations between absenteeism and other personnel statistics such as frequency of absenteeism events, age of an employee, sex and employee´s designation (blue collar or white collar). The study aims to analyze unique Human Resource Management (HRM) data set to find interdependences and visualize them by Multiple Correspondence Analysis (MCA) biplot and furtherly put the data into context of HRM decision processing.

MCA presents strong visualization technique detecting and representing underlying structures in a data set by expressing each group of categorical variable as a point in a low-dimensional Euclidean space. Nowadays it is possible to grasp enormous amounts of visualization and use more data and process it with visual pattern recognition as the basis of exploratory analysis. Visualization plays important role in HRM for the need of expressing the right things by the right way as described in [3].

Analyzed data are strongly oriented on absenteeism and present anonymized personnel information about more than five thousand employees from six production companies from Moravian-Silesian region including absenteeism in last two years 2015 - 2016. Absenteeism can generally include all events in which an employee is absent which includes holiday, sick leave and other absenteeism such as visiting doctor, blood donation, special events defined in Labor code etc. This study is oriented only on sick leave absenteeism because of growing trend of sick leave absenteeism in total absenteeism and becoming more and more costly feature for production companies as well as the fact that holiday and some special events defined in Labor code are unavoidable. Simple hypotheses were independently formulated by companies' management about sick leave absenteeism:

- older employees tend to have longer length of absenteeism events;
- younger employees tend to have more frequent and shorter absenteeism;
- white collars have lower absenteeism than blue collars;
- there are no differences in women and men absenteeism.

Some of the hypotheses are based on general observations but can be confirmed only through the proper research. Each of the hypotheses could be researched individually by using more accurate tool, but it is more optimal to compare them all together. Similar use of MCA in social-economic environment was described in [6], [9]

[1] Faculty of Economics VŠB-TUO, Sokolská třída 33, Ostrava, Czech Republic, ondrej.mikulec@vsb.cz

and [10]. We drew information from [2] to work with and correctly define HRM data. Stata11 is used to process the data and perform MCA.

Data Description

Data set represents summarization of the HRM data of over five thousand employees from companies coming from production area from Moravian-Silesian region. We define following sets of categorical variables, to research given hypotheses: cumulated length of absent work-days, amount of absenteeism events, sex, category and age. Similar description of HRM data for statistical processing can be found in [13].

Length of absenteeism (ABS) is based on the total amount of absent time throughout two past years. Five groups are defined: a group with zero absence during years 2015-2016, or in other words, people, which were not taking any sick days in last two years (0); second group was determined as those who were on a sick leave for less than two weeks (<2W) for which the employees are being paid by the company (except first three days of sick leave as per current legislation). Third group had been on a sick leave for more than two weeks but less than one month (2W-1M), which is still crucial short term absenteeism in which production companies are usually unable to replace sick employee by some replacement, i.e. agency worker. Fourth statistically important group had been on a sick leave for more than a month but less than three months (1M-3M) and the last group had the longest registered sick leave, longer than three months (3M+).

Second categorical variable was defined as amount of sick leave events during last two years. There had been a significant amount of employees with zero sick leave events (E_0) as well as with one sick leave event (E_1). Roughly ten percent of researched population had two registered sick leave events (E_2) and the rest was labeled as three or more sick leave events (E_3+). Furtherly we defined sex as Female (F) and male (M) gender.

Category of employee is represented by blue collars (BC) and white collars (WC) although by Czech legislation these categories are unofficial. We use this category differentiation to test whether there is a different association with absenteeism for workers physically involved in production process and office workers, management etc.

Last group was differentiated according to age. Employees were sorted to one of the preset groups according to their age: first group was defined as employees younger than 30 (A_<30), second group included employees older than 30 but younger than 40 (A_30-39), third group included employees older than 40 but younger than 50 (A_40-49), fourth group included those older than 50 but younger than 60 years old (A_50-59) and last group was set for all old as or older than 60 (A_60-65). Ten years age interval was set to cover employees in different life situations and possibly different association to sick leave absenteeism.

2 Principles of Multiple Correspondence Analysis

Correspondence analysis (CA) is an exploratory multivariate technique for graphical and numerical analysis of almost any data matrix with non-negative entries, but it principally involves tables of frequencies or counts. Because it is oriented toward categorical data, it can be used to analyze almost any type of tabular data after suitable data transformations, or recording [4]. MCA is an extension of CA, which allows one to analyze the pattern of relationships of several categorical dependent variables. Simple CA is primarily applicable to a two-way contingency table as in [12], leading to a map that visualizes the associations between two categorical variables. CA focus is on two-dimensional table of frequencies usually called contingency table and shows inner associations between two sets of categorical variables.

In MCA all variables have the same status unlike in CA. MCA tackles more general problem associations among a set of more than two categorical variables, which makes it more suitable for researching greater sets of data, especially with more than two sets of categorical variables. It represents a case of reminiscent of principal analysis, which involves all the cross tables of a set of categorical variables. By that, we obtain what is called a Burt matrix, i.e. in Table 1. Burt matrix is related quite simply to an indicator matrix as described in [5]. MCA is calculated as described in [14].

2.1 MCA Eigenvalue Correction

MCA codes data by creating several binary columns for each variable with the constraint that one and only one of the columns gets the value one. This coding schema creates artificial additional dimensions because one categorical variable is coded with several columns. Consequently, the inertia of the solution space is artificially inflated and therefore the percentage of inertia explained by the first dimension is severely underestimated. Formula with better estimation for eigenvalues according to [1] can be expressed as

$$\omega_J = \begin{cases} \left[\left(\frac{K}{K-1}\right)\left(\omega_i - \frac{1}{K}\right)\right]^2 & if\, \omega_i > \frac{1}{K} \\ 0 & if\, \omega_i \leq \frac{1}{K} \end{cases}, \tag{1}$$

where K is an amount of nominal variables, ω_i represents proportionally redistributed eigenvalues for each pattern of relationship, i stands for the given relationship, ω_j total eigenvalue and j for given pattern of relationship. Using this formula gives a better estimate of the inertia, extracted by each eigenvalue.

2.2 MCA procedure

MCA procedure represents a set of steps to perform the analysis and interpret the data. CA and MCA procedure is defined in six following steps as in [11], although the procedure is adjusted according to [8] to fit better for MCA.

1. define the objectives of MCA to assess associations among row and column categories and rows and columns themselves,
2. task formulation and creation of squared non-negative data matrix (Burt table),
3. fulfilling assumptions of compositional techniques (completeness of the input characters),
4. presentation of row, column or both categories in common chart where we look for suitable number of chart dimensions,
5. interpretation of the results by defining associated categories and comparing row and column categories,
6. verifying the results.

3 Interpreting MCA Results of HRM Data

MCA can be used also as exploratory analysis which does not require testing of predefined hypotheses and instead it can be used to identify patterns or hidden structure in data sets, which can suggest new research directions, but in this study we aim to analyze preset hypotheses. Data set is arranged to a form of Burt table and MCA is performed furtherly following MCA procedure from 2.2.

		ABS					EVENT				SEX		CAT		AGE				
		0	<2W	2W-1M	1M-3M	3M+	0	1	2	3+	F	M	BC	WC	<30	30-39	40-49	50-59	60-65
ABS	0	3214	0	0	0	0	3214	0	0	0	335	2879	2214	1000	218	317	1165	1290	224
	<2W	0	436	0	0	0	0	378	57	1	68	368	294	142	32	62	176	152	14
	2W-1M	0	0	430	0	0	0	298	105	27	66	364	355	75	16	60	158	166	30
	1M-3M	0	0	0	630	0	0	328	190	112	75	555	535	95	38	64	221	266	41
	3M+	0	0	0	0	456	0	171	173	112	50	406	408	48	17	36	98	242	63
EVENT	0	3214	0	0	0	0	3214	0	0	0	310	2684	2047	947	207	300	1091	1194	202
	1	0	378	298	328	171	0	1175	0	0	191	1204	1096	299	65	139	479	606	106
	2	0	57	105	190	173	0	0	525	0	59	466	448	77	31	55	163	230	46
	3+	0	1	27	112	112	0	0	0	252	34	218	215	37	18	45	85	86	18
SEX	F	335	68	66	75	50	310	191	59	34	594	0	280	314	18	37	203	310	26
	M	2879	368	364	555	406	2684	1204	466	218	0	4572	3526	1046	303	502	1615	1806	346
CAT	BC	2214	294	355	535	408	2047	1096	448	215	280	3526	3806	0	241	363	1337	1575	290
	WC	1000	142	75	95	48	947	299	77	37	314	1046	0	1360	80	176	481	541	82
AGE	<30	218	32	16	38	17	207	65	31	18	18	303	241	80	321	0	0	0	0
	30-39	317	62	60	64	36	300	139	55	45	37	502	363	176	0	539	0	0	0
	40-49	1165	176	158	221	98	1091	479	163	85	203	1615	1337	481	0	0	1818	0	0
	50-59	1290	152	166	266	242	1194	606	230	86	310	1806	1575	541	0	0	0	2116	0
	60-65	224	14	30	41	63	202	106	46	18	26	346	290	82	0	0	0	0	372

Table 1 Burt table

In Table 1 we can see amount of employees in each group in the diagonal and amount of employees from given group in observed group in other parts of the table. Large amount (3214 out of 5166) of employees have zero sick leave events. Only around 12% of observed employees (456) are females as data set represents productive companies with large part of physically working employees. Firstly was performed MCA for the whole data set including group of employees (in total 62%) with zero sick leave events during 2015-2016.

Number of observation = 5166

Dimension	Eigenvalue	Inertia	InertiaCum
dim 1	0.0586877	52.01	52.01
dim 2	0.0110916	9.83	61.84
dim 3	0.0024160	2.14	63.98
dim 4	0.0004438	0.39	64.37
dim 5	0.0000934	0.08	64.46
dim 6	0.0000528	0.05	64.5
Total	0.11284	100	

Table 2 MCA whole population results

MCA project row and column profiles into lower amount of dimensions. Cumulative inertia in Table 2 represents total amount of information displayed and captured by given amount of dimensions.

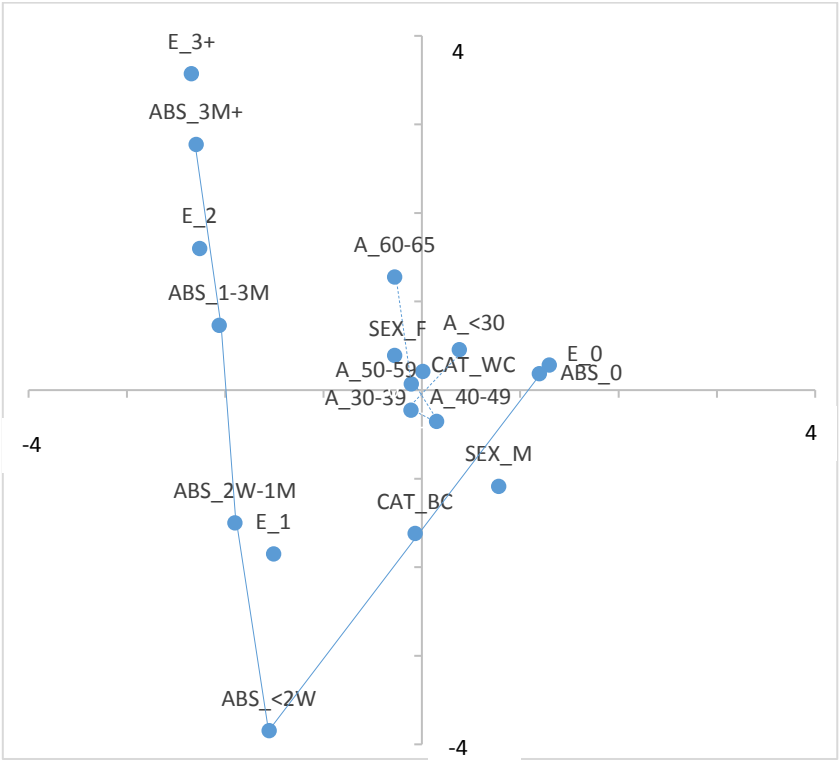


Figure 1 MCA whole population biplot

Common position of the categorical variables represented by points in biplot in Figure 1 shows generally low associations between length of absenteeism and other groups of categorical variables, although there is visible pattern with amount of absenteeism events showing that those, who are being absent more frequently tend to have also higher total length of absenteeism longer. This goes against expectations because frequent absenters were expected to spend short term sick leave per each event and thus not so much in total. Association or interdependence is generally defined between the points which are close to each other as the mass of one represented in the other one is significantly higher than in those who are further in the two dimensional biplot.

Number of observation = 1952

Dimension	Eigenvalue	Inertia	InertiaCum
dim 1	0.0192329	53.18	53.18
dim 2	0.0035622	9.85	63.03
dim 3	0.0008512	2.35	65.38
dim 4	0.0001716	0.47	65.85
dim 5	0.0001423	0.39	66.25
Total	0.0361678	100	

Table 3 MCA only ≥1 sick leave events results

Previous analysis required a correction to receive visually more synoptic results giving the management valuable information for decision processing. Following analysis was performed excluding the data of zero absenters to better fulfil the purpose of the study. By focusing only on those who had at least one sick leave event and excluding a large part of the data set we received more predicative results.

The amount of explained information in corrected model in Table 3 is higher which is expressed by increase in cumulative inertia compared to model including whole population in Table 2. The model is considered sufficient if the first dimension inertia is explained by more than 50% which is fulfilled in both models.

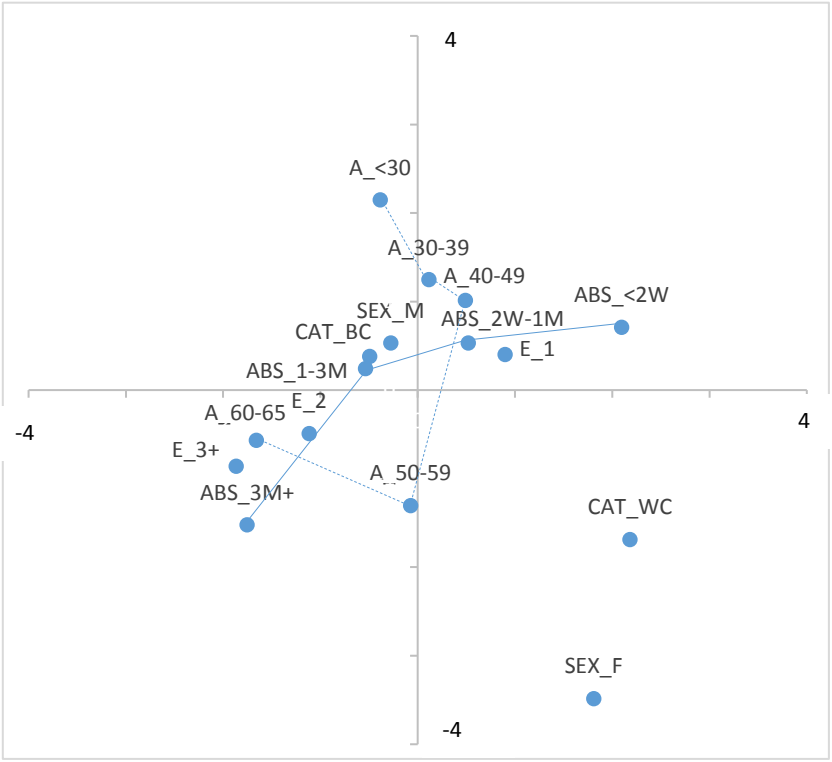


Figure 2 MCA only ≥1 sick leave events biplot

Corrected MCA in Table 3 showed both higher level of explained inertia and visually more synoptic results in Figure 2. In Figure 2 we can see connected dots representing age groups and also length of absenteeism. Pattern showing association between amount of sick leave events and total length of absenteeism observed in Figure 1 is visible also in Figure 2. We can conclude that higher age can be associated with both longer and more frequent absenteeism still rather explaining association based on percent occurrence, i.e. mass, and not causality. This fact does not correspond with the expectation that younger employees have more frequent absenteeism, although length of absence and age pattern confirm that younger employees tend to have shorter sick leave in total.

Expectation that WC tend to have lower absenteeism than BC is partially confirmed by the fact that BC are strongly associated with absenteeism 1 – 3 months in total length as well as being close to either 3 and more plus 2 weeks to 1 month. WC position shows no strong association with certain total length of sick leave, even though it is being closer to less than 2 weeks and from 2 weeks to 1 month absenteeism. Similar situation applies for the expectation that men are being equally sick as women but from opposite perspective. Men are more associated to certain absenteeism intervals while women are not associated to any particular total length of absenteeism.

4 Discussion

People analytics as data-driven approach to managing people at work are based on implementation of quantitative models into HRM decision processing and its influence on the company. Over the past decade, big data analytics and people analytics has been revolutionizing the way many companies do business. Common use for these techniques is described for recruitment, talent management, people retention etc. as well as market performance as in [7] and this study is oriented on interdependencies and associations of personnel data with focus on absenteeism and visualization of the results using MCA two dimensional biplot.

MCA was used to all together assess given hypotheses and there were presented two individual results in the study. One of the analyzed the whole population but had to be corrected due to fact that large part of observed

employees had zero sick leave events and thus strongly influenced MCA output as is visible in Figure 1. This was corrected by excluding this group and leaving only those employees with one or more sick leave events in order to present visually acceptable result. Also there is high probability that some of the sick leave was covered by holiday leave and thus have not appeared in the model. This way of dealing with illnesses is known mostly among WC employees because first three days of each sick leave are unpaid by employer according to Labor code. Nevertheless, used data set with so large researched population should avoid to be influenced by this.

Some of the hypotheses were confirmed such as that higher age is generally associated with higher total length of absenteeism and that WC tend to have lower sick leave than BC as discussed above. Some hypotheses were partially disapproved, such as assumption that younger employees tend to have higher sick leave frequency with lower absenteeism. Based on the results we conclude that MCA represents a suitable statistical tool to uncover interdependencies and associations for HRM data. HRM decisions supported by the results of the analysis can be used to various HR problems, such as directing of sanitation residence, vitamin packages, active discussion about sick leave absenteeism with subordinates after sick leave return, prevention for unnecessary absenteeism etc., generally focusing on specific group rather than allocating time and effort at the same rate for each group. Also well visualized result works as a great general overview of employees' behavior and synoptical distribution of variously differentiated groups in the company.

Acknowledgements

The article was written thanks to the support of grant SGS2017/148 and grant Ostrava stipendist 2016/2017.

References

[1] Abdi H., Valentin D.: Multiple Correspondence Analysis. *Encyclopedia of Measurement and Statistics.* Thousand Oaks (CA): Sage, 2007.
[2] Armstrong, M., Taylor, S.: *Armstrong's Handbook of Human Resource Management Practice*. Kogan Page; 13 ed., 2014.
[3] Few, S.: Data Visualization for Human Perception. *The Encyclopedia of Human-Computer Interaction*, 2nd Ed. 2015.
[4] Greenacre, M. J.: *Correspondence analysis in practice*. London: Academic Press, 1993.
[5] Greenacre, M. J.: Interpreting multiple correspondence analysis. *Applied Stochastic Models and Data Analysis* vol. 7, iss 2, 1991, 195–210.
[6] Guinot, C., Latreille, J., Malvy, D., Preziosi, P., Galan, P., Hercberg, S., Tenenhaus, M.: Use of multiple correspondence analysis and cluster analysis to study dietary behaviour: Food consumption questionnaire in the SU.VI.MAX. cohort. *European Journal of Epidemiology* vol. 17, iss. 6, 2001, 505–516.
[7] Harel, G. H., Tzafrir, S. S.: The effect of human resource management practices on the perceptions of organizational and market performance of the firm. *Human Resource Management* vol. 38, iss. 3, 1999, 185–199.
[8] Hoffman, D. L., Leeuw D. J: Interpreting multiple correspondence analysis as a multidimensional scaling method. *Marketing Letters* vol. 3, iss. 3, 1992, 259-272.
[9] Kaciak, E., Louviere, J.: Multiple Correspondence Analysis of Multiple Choice Experiment Data. *Journal of Marketing Research* vol. 27, no. 4, 1990, 455-465.
[10] Kalayci, N., Basaran, M. A.: Combined approach using multiple correspondence analysis and log-linear models for student perception in quality in higher education. *Procedia Economics and Finance*, vol. 17, 2014, 55-62.
[11] Meloun, M., Militký, J., Hill, M.: *Statistická analýza vícerozměrných dat v příkladech*, Praha: Academia, 2 ed., 2012.
[12] Mori, I., Kuroda, M., Makino, N.: Multiple Correspondence Analysis. *SpringerBriefs in Statistics: Nonlinear Principal Component Analysis and Its Applications*, 2016, 21-28.
[13] Noe, R., Hollenbeck, J., Gerhart, B., Wright, P., *Human resource management: Gaining a competitive advantage*. McGraw-Hill Education, 9 ed., 2014.
[14] Roux, L. B.: Multiple Correspondence Analysis (Quantitative Applications in the Social Sciences). *Quantitative Applications in T*, 2009.

The efficiency of stocks investment strategy with the use of chosen measures of deterministic chaos to building optimal portfolios

Monika Miśkiewicz-Nawrocka[1], Katarzyna Zeug-Żebro[2]

Abstract. In recent years, in addition to classical methods of portfolio analysis have been developed tools that are both modifications of these concept as well as new, alternative diversification techniques of investment portfolio which take into account for example the indicators of fundamental analysis. A new approach proposed in the paper is the use of the measures for identifying chaos, i.e. the largest Lyapunov exponent and the Hurst exponent. Since determinism of chaotic time series indicates on potential possibility of their prediction, it is also expected that has a significant impact on the construction of optimal portfolio.
The paper aims to construct optimal portfolios determined based on the largest Lyapunov exponent and the Hurst exponents. The test will be conducted on the basis of the financial time series in 2005-2015.
Keywords: portfolio analysis, largest Lyapunov exponent, Hurst exponent.

JEL Classification: C3, C8, G11, E4
AMS Classification: 91B28

1 Introduction

The construction of an optimal portfolio proposed by H. Markowitz [6,7] started intensive development of scientific field which is a portfolio analysis. Research conducted for many years in various scientific centers have provided new tools and approaches for estimating the shares in the optimal portfolio, e.g. taxonomic measure of investment attractiveness (TMAI) [14,15], value investing, growth investing, dividend investing [2, 16], innovation investing [13]. A new approach proposed by the authors is the use to construct the optimal portfolio of two measures for identifying chaos, i.e. the largest Lyapunov exponent and the Hurst exponent [9,10, 11, 12].

The aim of the paper will be an attempt to diversify the risk of the investment portfolio, based on the constructed optimal portfolios determined on the value of the largest Lyapunov exponent and Hurst exponent. In the study we used the financial time series of the WIG20 index, which at the time of portfolios building were listed on the Warsaw Stock Exchange for at least 10 years. The WIG20 is the stocks index of 20 largest companies listed on the Warsaw Stock Exchange, updated at the end of each quarter. The data cover a period from 1.01.1994 to 31.12.2016, and stocks portfolios were built at the end of each year between 2004 and 2015.

2 The largest Lyapunov exponent

The pair (X, f), where X is the state space and $f : X \to X$, $X \subset R^m$, is called the dynamic system with discrete time (cascade)[12]. These dynamic systems can be described by first order recurrence equation of the form:

$$x_{t+1} = f(x_t), \qquad t = 0,1,2,\ldots, \tag{1}$$

where x_t, x_{t+1} - state on the system respectively at the moments t and $t+1$.

Lyapunov exponents are defined as limits [17]

[1] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, monika.miskiewicz@ue.katowice.pl.
[2] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, katarzyna.zeug-zebro@ue.katowice.pl.

$$\lambda_i(x_0) = \lim_{n\to\infty} \frac{1}{n} \ln\left|\mu_i(n, x_0)\right|, \qquad i = 1. \ldots. m. \quad \text{for } m \geq 1, \tag{2}$$

where $\mu_i(n, x_0)$ are the eigenvalues of the Jacobi matrix of mapping f^n. f^n is an *n*-fold submission of function *f*, and *f* is the function that generates a dynamic system.

The Lyapunov exponents measure the rate of divergence or convergence of neighboring trajectories, i.e. the level of chaos in a dynamic system. The largest (maximal) Lyapunov exponent allows to specify the extent of a change (an increase or a decrease) in the distance between the current state x_N of the system and its nearest neighbor x_i in the evolution of the system, and also estimate the distance between the vectors x_{N+1} and x_{i+1}. Based on this distance the value of the forecasts $\hat{x}_{N+1}$ is determined [3,9].

For real-time series, if you do not know a generator function *f*, the largest (maximal) Lyapunov exponent is estimated based on the relation [15]:

$$\Delta_n = \Delta_0 \cdot e^{n\lambda_{\max}}, \tag{3}$$

as the direction component of the regression equation [4, 5, 11]:

$$\ln \Delta_n = \ln \Delta_0 + \lambda_{\max} n, \tag{4}$$

where Δ_0 is the initial distance between two initially close (in the Euclidean distance sense) points of the reconstructed state space, Δ_n is the distance between these points after *n* iterations and $\lambda_{\max}$ is the largest (maximal) Lyapunov exponent.

Consider a one-dimensional time series, composed of *N* observation $(x_1, x_2, \ldots, x_N)$. Of all the vectors x_t^d of reconstructed state space we choose the vector closest to the vector x_N^d (in terms of Euclidean distance) and it is denoted by $x_{\min}^d$. Let $\Delta_{\min}$ denote the distance between x_N^d and $x_{\min}^d$, and Δ_1 - the distance between x_{N+1}^d and $x_{\min+1}^d$. Assuming that $\Delta_1/\Delta_{\min}$ is a small change in the evolution of the system, the distance between vectors x_{N+1}^d and $x_{\min+1}^d$ is given by [3]:

$$\Delta_1 \approx \Delta_{\min} \cdot e^{\lambda_{\max}}, \tag{5}$$

where $\lambda_{\max}$ is the largest (maximal) Lyapunov exponent.

3 Hurst exponent

The Hurst exponent [4] is another measure that allows for the classification of time series, i.e. to distinguish chaotic time series generated by deterministic dynamic systems from the stochastic time series. The exponent has a value in the range $\langle 0,1 \rangle$. If the time series is generated by a random walk (or a Brownian motion process) it has the value of $H = 0.5$. If $0 \leqslant H < 0.5$ the time series is antipersistent or ergodic. For a series, for which $0.5 < H \leqslant 1$, the series is persistent, i.e. reinforcing trend.

One of the methods of calculating the Hurst exponent is the method of the rescaled range R/S. For time series $\{x_1, x_2, \ldots, x_N\}$ it runs through the following steps[1, 12, 18]:

(1) transform the above time series into $m = N - 1$ logarithmic rates of return:

$$y_k = \log(x_{k+1}/x_k), \quad k = 1. 2. \ldots. N-1. \tag{6}$$

(2) Next, share a series (6) on *T* parts made up of *t* elements: $T = [m/t]$, where [ ] denotes the integer part of the argument. If the quotient m/t is not an integer then $tT < m$ and we use values y_k for $k = 1. 2. \ldots. tT$.

(3) In the next step, define the:

$$z_{ij} = y_{ij} - \bar{y}_j, \tag{7}$$

where: y_{ij} is the *j*-th value in the *i*-th interval and $\bar{y}_j = \frac{1}{t}\sum_{i=1}^{t} y_{ij}$.

(4) A sequence of partial sums z_{ij} for each *i*, is given by:

$$q_{ij} = \sum_{l=1}^{i} z_{lj}, \qquad i = 1. 2. \ldots. t, \quad j = 1. 2. \ldots. T. \tag{8}$$

(5) The range of the *i*-th interval is defined as:

$$R_j = \max(q_{ij}) - \min(q_{ij}). \tag{9}$$

(6) Calculate the rescaled range series (*R*/*S*):

$$\alpha_{jt} = R_j / S_j, \tag{10}$$

where: $S_j = \sqrt{\frac{1}{t}\sum_{i=1}^{t} z_{ij}^2}$.

(7) Next, calculate:

$$(R/S)_t = (1/T)\sum_{j=1}^{T} \alpha_{jt}. \tag{11}$$

(8) The above procedure is carried out for different lengths of time series *t*.

(9) Finally, the value of the Hurst exponent is the slope of the graph of the logarithms $(R/S)_t$ to the axis of logarithms *t*.

4 The optimal stock portfolio

Optimal allocation of shares in portfolio is issued by solving the following optimization problems:

- Model I

$$\max\left(\sum_{i=1}^{m} \lambda_{\max i} x_i\right)$$
$$R_p \geq R_0$$
$$\sum_{i=1}^{m} S_i x_i \leq S_0$$
$$\sum_{i=1}^{m} x_i = 1$$
$$0.4 \geq x_i \geq 0, \quad i = 1,\ldots,m$$

- Model II

$$\max\left(\sum_{i=1}^{m} H_i x_i\right)$$
$$R_p \geq R_0$$
$$\sum_{i=1}^{m} S_i x_i \leq S_0$$
$$\sum_{i=1}^{m} x_i = 1$$
$$0.4 \geq x_i \geq 0, \quad i = 1,\ldots,m$$

where:

S_i - the standard deviation for *i*-company,

x_i - contribution of *i*-share in the portfolio,

R_p - the expected rate of return on the portfolio: $R_p = \sum_{i=1}^{m} x_i R_i$,

R_i - the expected rate of return for *i*-company,

R_0 - average rate of return for companies,

S_0 - mean standard deviation,

$\lambda_{\max i}$ - the largest (maximal) Lyapunov exponent for *i*-company,

H_i - the Hurst exponent for *i*-company.

5 The purpose and conduct of the study

In the study we used financial time series of the WIG20 index, which at the time of portfolios building were listed on the Warsaw Stock Exchange for at least 10 years – from 2004 to 2015. The optimal portfolios were built at the end of each year between 2004-2015.

The largest Lyapunov exponent and the Hurst exponent for the analyzed companies was estimated on the basis of the algorithms described in points 1 and 2. For this were used financial time series which were set up with logarithms of daily returns of closing price indexes of selected companies in period from 1.01.1994 to 31.12.2015[3]. The value of the largest Lyapunov exponent allowed to choose the companies, whose time series are characterized by chaotic dynamics i.e. the largest Lyapunov exponent is positive and statistically significant (R^2>0,3).

In the next stage of the studies it was constructed 24 investment portfolios based on solving optimization problems: model I and model II. The optimal portfolios based on model I conclude only companies with positive and statistically significant largest Lyapunov exponent. Table 1 shows the value of the expected rate of return and risk of constructed portfolios.

		Portfolio 2004		**Portfolio 2005**		**Portfolio 2006**		**Portfolio 2007**		**Portfolio 2008**		**Portfolio 2009**	
Model I		BPH BZW	- 1.0000	-	-	BPH	1.0000	BZW	1.0000	ACP BZW CERS KGHM MBK	0.4000 - 0.4000 - 0.2000	BZW CERS MBK PEO PKN	0.0807 0.4000 - 0.1193 0.4000
I	**Rate of return**	-0.00046		-		-0.00054		-0.00012		0.00318		-0.00035	
	Risk	0.000223				0.000425		0.000646		0.000361		0.000389	
		Portfolio 2004		**Portfolio 2005**		**Portfolio 2006**		**Portfolio 2007**		**Portfolio 2008**		**Portfolio 2009**	
Model II		BPH BZW DBC MBK	0.2000 0.4000 0.4000 -	BPH BZW MBK	0.3537 0.4000 0.2463	BPH BZW KTY MBK	0.2807 - 0.4000 0.3193	BPH BZW KGHM MBK	0.247061 0.36304 0.3899 -	ACP BZW CERS KGHM MBK PEO PND TPSA	0.4000 0.1991 - - - - 0.4000 0.0009	ACP BZW CERS CEZ KGHM MBK PEO PKN TPSA	0.4000 - 0.4000 0.1092 0.0908 - - - -
II	**Rate of return**	0.00018		-0.00135		-0.00161		0.00636		0.00423		-0.00072	
	Risk	0.000109		0.000116		0.000147		0.011502		0.000366		0.000362	
		Portfolio 2010		**Portfolio 2011**		**Portfolio 2012**		**Portfolio 2013**		**Portfolio 2014**		**Portfolio 2015**	
Model I		BZW MBK PXM	0.2426 0.3574 0.4000	BHW KGHM PEO	0.4000 0.2000 0.4000	ACP BHW BRS KGHM MBK PEO PKN	0.4000 - 0.3198 - 0.2802 - -	ACP BHW KGHM MBK PEO	0.4000 - 0.0470 0.1530 0.4000	BZW KGHM LPP MBK PEO PKN SNS	- -- 0.2000 0.4000 - - 0.4000	ACP BZW MBK PEO PKN	0.4000 0.4000 - 0.2000 -
I	**Rate of return**	-0.00024		0.00085		-0.00013		-0.00017		0.00062		0.00023	
	Risk	0.000118		0.000171		0.000117		0.000109		0.000100		0.000108	
		Portfolio 2010		**Portfolio 2011**		**Portfolio 2012**		**Portfolio 2013**		**Portfolio 2014**		**Portfolio 2015**	
Model II		ACP BZW CEZ KGHM MBK PEO PKN PXM TPSA	0.4000 0.4000 0.0107 - 0.0785 - - - 0.1108	ACP BHW KGHM MBK PEO PKN TPSA	0.4000 0.4000 - 0.1905 0.0095 - -	ACP BHW BRS KGHM MBK PEO PKN TPSA	0.1113 0.3012 0.1875 - 0.4000 - - -	ACP BHW BZW KGHM MBK PEO PKN	0.1402 - 0.4000 0.3853 0.0744 - -	ACP BZW KGHM LPP MBK PEO PKN PKO	- 0.2214 0.3786 - 0.4000 - - -	ACP BZW CCC EURO KGHM LPP LTS MBK PEO PGN PKN PKO	- -- - - - - 0.4000 0.2665 - - 0.3335 -
II	**Rate of return**	0.11081		0.00102		-0.00082		-0.00018		0.00012		0.33380	
	Risk	0.000112		0.000342		0.000142		0.000201		0.000156		0.000138	

Table 1 The expected rates of return. risk and compositions of constructed portfolios

[3] The data comes from stooq.com. Date of access: 12.29.2009, 2.12.2010, 10.30.2012, 4.24.2017.

In 2005, the largest Lyapunov exponent for analyzed time series was either negative or statistically insignificant. Thus, an optimal portfolio wasn't built.

On the basis of the data presented in Table 1 we can conclude that for the optimization model II the most of portfolios are characterized by higher expected rate of return than for the optimization model I portfolios. The highest level of expected return was characterized for the optimization model II portfolio in 2015 at one of the lowest risk level.

Table 2 shows the obtained annual rates of return for the designated portfolios and for the optimization models.

Model	Portfolio 2004	Portfolio 2005	Portfolio 2006	Portfolio 2007	Portfolio 2008	Portfolio 2009
I	0.45898	-	-0.00162	-0.55032	0.31700	0.12449
II	0.36006	0.36246	0.08476	-0.60455	0.46449	-0.04200
Model	**Portfolio 2010**	**Portfolio 2011**	**Portfolio 2012**	**Portfolio 2013**	**Portfolio 2014**	**Portfolio 2015**
I	-0.26466	0.47229	0.49083	0.05065	-0.24589	0.10431
II	0.01485	0.30372	0.49248	-0.00139	-0.34130	0.14325

Table 2 The annual rate of return for constructed investment portfolios

Analyzing the obtained rates of return for the designated portfolios (Table 2). it can be seen that the largest profit would give the investment in portfolio 2012 and portfolio 2008 in model II. and then in portfolio 2004 and portfolio 2006 in model II. The lowest rate of return obtained for the portfolio 2007 in both optimization models.

Figure 1 compares the rates of return of obtained portfolios with the return of the WIG 20 index for the years 2004-2015.

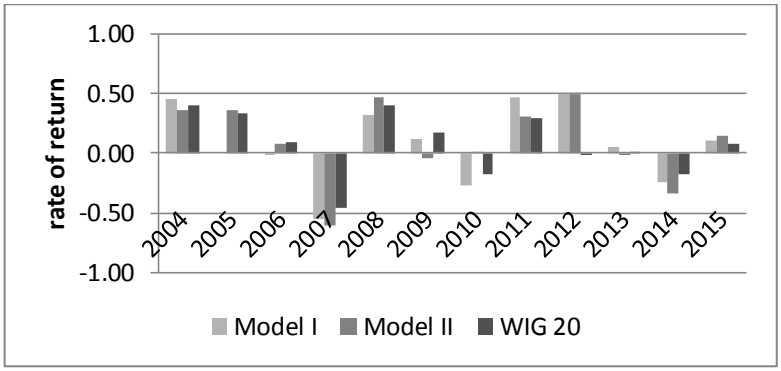


Figure 1 The rate of return of portfolios and index WIG 20

On the basis of the data in Fig. 1 we can conclude that most portfolios in the optimization model I or model II are characterized by higher rate of return than index WIG 20. The exception are years 2006. 2007. 2009 and 2014. The calculated 10-year return rates for the analyzed strategies showed that the most profitable strategy was to invest in the WIG20 index (0.6813). then the II model (0.6561). The worst long-term strategy was model I (0.3699).

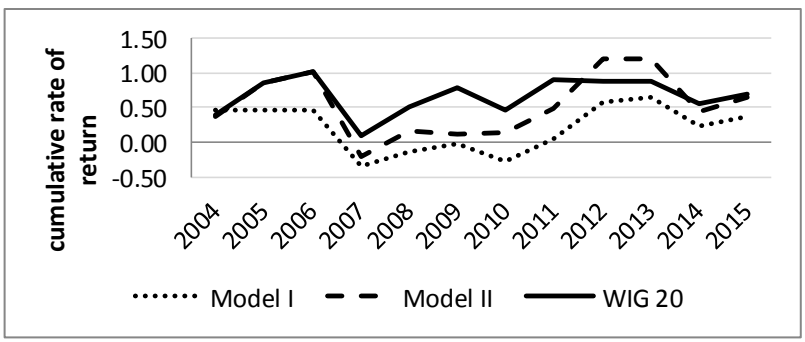


Figure 2 The cumulative rate of return of portfolios and index WIG20

Figure 2 presents the cumulative rate of return of portfolios and index WIG20. Obtained results (fig. 2) show also that proposed in the paper approach gives generally better results during bull market (2005-2007 and 2009-2014) then bear (2008. 2015).

7 Conclusions

In the paper was presented an attempt to construct the optimal portfolio of shares based on the value of the largest Lyapunov exponent and the Hurst exponent. The study has revealed that the portfolios built on the basis of the largest Lyapunov exponent and Hurst exponents give similar results due to the annual rate of return to the index WIG20 in 2005-2015. Obtained results show also that proposed in the paper approach gives generally better results during bull market (2005-2007 and 2009-2014) then bear (2008. 2015).

References

[1] Chun. S.H.. Kim. K.J.. Kim. S.H.: *Chaotic analysis of predictability versus knowledge discovery techniques: case study of the Polish stock market*. Expert Systems. Vol. 19. No. 5. 2002. 264 – 272.

[2] Damodaran. A.: *Investment philosophies. Successful strategies and the investors who made them work*. New Jersey: John Wiley & Sons. Inc.. 2012.

[3] Guégan. D.. Leroux. J.: *Forecasting chaotic systems: The role of local Lyapunov exponents*. Chaos. Solitons & Fractals. vol. 41. 2009. 2401 – 2404.
http://mme2016.tul.cz/conferenceproceedings/mme2016_conference_proceedings.pdf

[4] Hurst. H. E.:. *Long term storage capacity of reservoirs*. Trans. Am. Soc. Eng. 116. 1951. 770–799.

[5] Kantz. H.. Schreiber. T.: *Nonlinear time series analysis*. Cambridge University Press. Cambridge. 1997.

[6] Markowitz. H.: *Portfolio Selection*. Journal of Finance. 1952. 77-91.

[7] Markowitz. H.: *Portfolio Selection: Efficient diversification of investments*. 94. Cowles Foundation. New Haven. CT. 1952.

[8] Mastalerz-Kodzis. A.. Pośpiech. E.: *Fundamental and Behavioral Methods in Investment Decision Making*. in: Financial Management of Firms and Financial Institutions. Wydawnictwo Technicznego Uniwersytetu w Ostrawie. Ostrawa. 2011. 250-257.

[9] Miśkiewicz-Nawrocka M.. Zeug-Żebro K. (2016): *The application of chosen measures of deterministic chaos to building optimal portfolios*. [In:] A. Kocourek. M. Vavroušek (eds.). *Proceedings of 34th International Conference Mathematical Methods in Economics*. Technical University of Liberec. Liberec. Faculty of Science. Czech Republic. 2016. 570-575.

[10] Miśkiewicz-Nawrocka. M. (2016). *Efektywność strategii inwestycji w akcje spółek notowanych na GPW w Warszawie z wykorzystaniem największego wykładnika Lapunowa oraz wykładnika Hursta*. [in:] J. Mika. M. Miśkiewicz-Nawrocka (red.) *Metody i modele analiz ilościowych w ekonomii i zarządzaniu*. cz.8. Wydawnictwo UE w Katowicach. Katowice. s. 46-57.

[11] Miśkiewicz-Nawrocka. M.. Zeug-Żebro. K.: *The application of Lyapunov exponents to building optimal portfolios*. Studia Ekonomiczne 221. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach. Katowice. 2015. 61-72.

[12] Miśkiewicz-Nawrocka. M.: *Zastosowanie wykładników Lapunowa do analizy ekonomicznych szeregów czasowych*. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach. Katowice. 2012.

[13] Nawrocki. T.. Jabłoński. B.: *Inwestowanie na Rynku Akcji. Jak Ocenić Potencjał Rozwojowy Firm Notowanych na GPW w Warszawie*. Wydawnictwo CeDeWu. 2011.

[14] Rosenstein. M. T.. Collins. J. J.. De Luca. C. J.: *A practical method for calculating largest Lyapunov exponents from small data sets*. Physica D. vol. 65. 1993. 117 – 134.

[15] Tarczyński. W.: *A taxonomic measure of the attractiveness of investments in securities*. Przegląd Statystyczny No 3/94. 1994. 275–300.

[16] Tarczyński. W.: *The fundamental attitude to building a stock portfolio*. Argumenta Oeconomica No 1(7). Wrocław. 1999. 153–168.

[17] Zaremba. A.: *Giełda. Skuteczne strategie nie tylko dla początkujących*. Helion. Gliwice. 2013.

[18] Zawadzki. H.: *Chaotyczne systemy dynamiczne*. Akademia Ekonomiczna. Katowice. 1996.

Research on probabilistic risk evaluation of business system development Project Based on Requirements Analysis

Shinji Mochida[1]

Abstract. The term "business system" refers to the methods which use to manage their information or knowledge. All companies employ business systems during the development of new software, mechanical equipment or other projects. Additionally In most business system development projects are planed in order to reduce amount of running cost or increase benefit. Thus More efficient project management is needed in order to maintain schedules and reduce cost. Although there are several stages in business system development projects. There are many stages of planning, development, system examination, and upgrade. And there are some factors that cause over cost or schedule delay of projects. Especially in early stage there are risks of projects. In business system development projects, there are stages of planning, development, system examination, and upgrade. Additionally In most business system development, costs and completion dates are difficult to predict. More efficient project management is needed in order to maintain schedules and reduce cost. Narrowing down requirements to applicable specifications is effective means to shorten schedules and reduce costs. Also evaluating risk properly avoids schedule delay and over cost. This paper aims to evaluate risk of business system development project using requirement analysis with cost share rates and probabilistic risk evaluation. Cost share rate is defined as the percentage of total cost assigned to each requirement. Cost share rate could indicates essentiality of requirements. Requirements need large cost share rate must have large risk,should be under strict control. At the conclusion, this paper shows the potential to evaluate risk of business system development using cost share rate and probabilistic risk evaluation. In the future, this method would be able to help evaluate risk properly in business system development projects in order to shorten schedule and reduce costs.

Keywords: Cost Share Rate, Project Management, β distribution, Bayesian analysis, Expected monetary value, Risk management

JEL classification: C11, O22
AMS classification: 91B06

1 Introduction

Business System development projects are challenging in that are many requirements demanded from customers even while these requirements are proposed with the same priority. Thus it is important to narrow down and prioritize requirements according to their essentiality and criticality to finish on schedule. Although system developers estimate according to the complexity of projects [2], but customers expect the cost to be based on the number of requirements they propose. Then customers and system developers estimate differently, there are often **conflicting estimates**. Thus, this paper proposes cost share rate for business system development projects based on requirements analysis in order to estimate accurately. Cost share rate is defined as the percentage of total cost assigned to each requirement.Cost share rate could distinguish essential requirements. And requirements that has large cost share rate must have large risk, also should be under strict control. Because big change or modification for essential requirement give large impact to costs or schedule. Then this paper aim to propose a method to identify essential requirements in order to estimate accurately. Then risks are assigned to each essential requirements with cost share rate and probability. This research show methods in following steps; (see Figure 1). First, this research predict cost for requirement version one (see Table 5). Next this research compares with predicted estimate and total cost at completion of the past project (see Table 6).

2 Previous Studies

One purpose of this paper is showing that risk management could contribute reducing cost of business system development project. Additional this paper demonstrates the potential to evaluate risk by requirements analysis for business system development project management. This research is not focused on estimating costs based on

[1]University of Marketing and Distribution Sciences/Faculty of Commerce, 3-1 Gakuen Nishimachi Nishiku Kobe Japan, mochida@red.umds.ac.jp

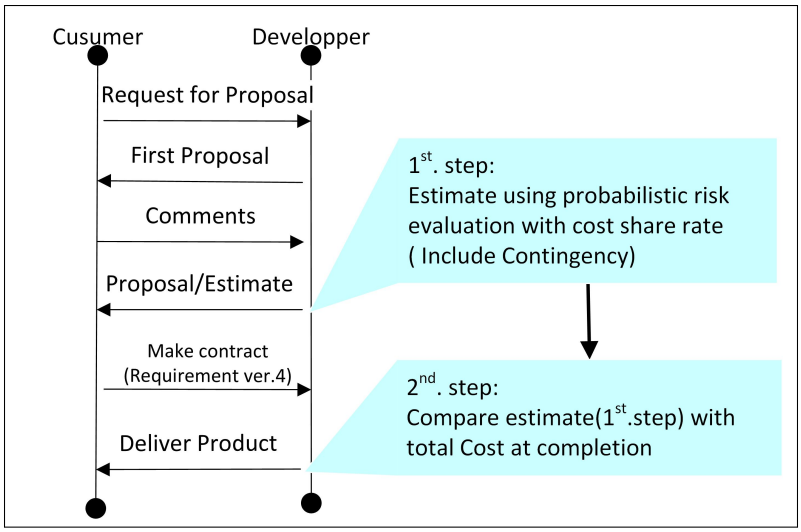


Figure 1 Steps of business system development project with customer and developer.

the method to measure amount of source code of system, but rather allocating costs to each requirement. Previous research typically focused on either schedule adherence,cost estimation or productivity. Improving productivity contribute to finish project fine [3]. About changing requirement in mechanical engineering design, one example of requirements analysis research explored the ability of predicting requirements change through graphical models of the requirements documents and historical change trends [4]. Again, this evolution of requirements analysis has not been used to distribute project costs to the individual requirements or to prioritize the requirements.

3 Project Management for Business System development

3.1 Project management and Risk management

In project management there are two important methods [5]. One is schedule management, and another one is risk management.On the other hand Equation 1 shows there are controllable factors or uncontrollable factors in business. Sale is uncontrollable factor, because sale come under the influence of markets, customers. Although cost is controllable factor, cost includes payment, material costs, for example, payment for staff could be cut off by manager. It is important to identify which factor is controllable and uncontrollable. And how appropriately controllable factor could be controlled. Also there is possibility uncontrollable factor could be controlled with Bayesian analyses. Usually extra budget is settled aside for refinement or fixing trouble in project management. This extra budget is called for contingency budget or only contingency. If risk management would work well, contingency budget would not be used, then contingency budget would come to benefit. At the result prospect of profit would increase. Thus risk management could contribute reducing contingency cost, and risk management has potential to increase contingency profit.

$$Gaining = Sale - Cost \tag{1}$$

3.2 Risk analysis

Risk is defined as factors that make uncertain when they will achieve their objectives under ISO31000. Usually risk Analysis is started from risk identification in risk management. Then, risk evaluation is considered by qualitative evaluation and quantitative evaluation. Thus quantitative risk analysis is calculated by possibility $\times$ cost. This calculated risk(cost) should be spend ,if risks comes up.It is called expected monetary value. Risks is evaluated by expected monetary value Equation 2. And risks are prioritized by the order of expected monetary value. In quantitative evaluation for risk management usually probability is given subjectively by staffs subjectively. Or probability is given by experts, Delphi method or questionnaire for skilled staffs.Thus getting accurate probability is very essential to cucullate expected monetary value correctly. Thus this paper takes into account β distribution to calculate probability,and expect monetary value. This paper propose conditional possibility in order to cucullate risk correctly.In this research conditional possibility is given to essential risks. In many project risk comes up by misunderstanding requirements or defects in requirements from clients in business system development project. It is vital to get certain requirements and predict risk in requirements properly.

$$Risk(Expected\,Monetary\,Cost) = Probability \cdot Cost \tag{2}$$

4 Cost Prediction Methods for Business System Development

4.1 Cost Prediction Methods

Proper estimate is essential to finish projects on schedule and under budget. Over cost or schedule delay is caused by missing estimate. Usually amount of program source code is predicted by some prediction method in order to estimate in business system development project. Then amount of program source code is converted into base monetary cost. Next total cost is made by adding contingency cost to base monetary cost. It is finally budget for project. Then this section explains some current methods to predict costs of business system development projects. There is typical methods to estimate for business system,typical methods; COCOMO method and Function Point method. Both methods predict costs for business system. In case of COCOMO method, it estimate by amount of program source code.On the other hand In case of Function Point method it accumulates points according to the complexity of system;the number of db tables,dialog boxes, print forms and interfaces. Acquired points could be converted into cost.

4.2 COCOMO Method

COCOMO Method, [6], estimate the duration (Person-Months) in system development projects. In the COCOMO method, volume of source code is estimated by equation 3. Duration (Person-Months) could be calculated with dividing volume of source code by the number of staffs. COCOMO Method propose equation 3. And it uses the parameters as follows:

C_e : estimate duration (PM:Person-Months) for expectation
C_0 : estimate volume of source code
P_1 : parameter for estimated productivity
P_2 : exponent parameter for software development
P_3 : calibration parameter

A challenge with this method is the parameters used for the cost estimation method are empirically derived and contextually dependent on many different factors, such as team size, project complexity, cultural environment, and others.

$$C_e = (C_0 \cdot P_1)^{P_2} \cdot P_3 \tag{3}$$

4.3 Function Point Method

Function Point Method estimate duration (PM: Person-Months) as those of COCOMO methods. In the Function point method it is necessary to count the number of internal and external files, tables and internal and external interfaces. Function Point Method propose equation 4. And it uses the parameters as follows:

C_e : estimate duration (PM: Person-Months) for expectation
F_p : estimate function points
F_1 : function points
P_1 : parameter for estimated productivity
P_2 : calibration parameter

$$C_e = (F_p \cdot P_1),\ F_p = F_1 \cdot P_2 \tag{4}$$

4.4 Other Methods to Estimate

Additionally, there is another cost estimate method as Experience method.In the experience method, total cost is estimated based on previous experiences. In these methods, there are often gaps between system developers cost estimates and customer expectations. This results from differences in how developers and customers group costs. Estimating correctly is important to finish building system on schedule and under budget. Thus, translating requirements into factors to estimate is essential. Certain requirements are needed to estimate properly. But there is no cost estimate method taking certainty of requirements into account.This paper considers taking certainty of requirements into account to estimate in order to gain customer agreement.

5 Requirement Analysis using Linguistic analysis

Taking correct requirements is essential to estimate properly.And over cost or schedule delay is caused by missing evaluation of requirements. Otherwise over cost or schedule delay is caused by many remediation of requirements. Remediation for essential requirement and uncertain requirement has large risk.thus It is vital to distinguish the requirement which gives large impacts to specification or budget of project.This research propose the method to distinguish influential requirements that has large risks.This paper shows a method to distinguish an influential requirements with with linguistic analysis and cost share rate. Cost share rate is defined as the percentage of total

cost assigned to each requirement. This paper analyzes the requirements that were requested in past small system development projects by linguistic analysis. This small project is building a knowledge collecting system. In this project requirements were revised four times. Thus this paper analyzes requirement version one and version four. This paper predict risks from the result from analyzing requirement version one. Risk is considered as cost in risk management. This results;acquired cost is compared with the actual cost at the completion of project. If risk for requirement could be predicted properly, it would contribute project management. In this research,overlapping keywords are extracted form each requirements with linguistic analysis.Overlapping keywords are words that appear in one requirement and groups of keywords that appear in each category or phase in system development project. Categories or phases are Design, Development, Print, Test, Interface and Document. Overlapping keywords indicates relationships between one requirement and each other. The number of relationships that each

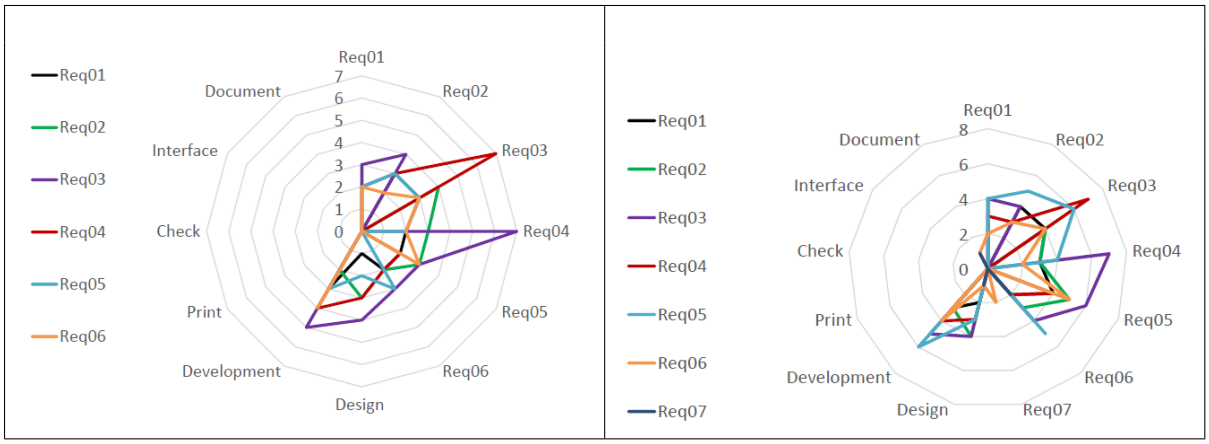


Figure 2 Left panel: Extract keywords from requirement version one and Relation with typical category. Right panel: Extract keywords from requirement version four and Relation with typical category.

requirement has with other requirements indicate essentiality and importance. Steps of linguistic analysis are as follows:

1. Extract keywords from each requirement in version one. And count overlapping key words from extracted keywords, also count overlapping key words from extracted keywords and group of typical keywords of each category (see Figure 2).
2. Extract keywords from each requirement in version four. And count overlapping key words from extracted keywords, also count overlapping key words from extracted keywords and group of typical keywords of each category (see Figure 2).
3. Distinguish essentiality of each requirements from these results, and distinguish categories that each requirements belong in (see Table 1).

6 Prediction of Cost Share Rate

Usually costs for system development projects are estimated by grouping costs with the number of dialog boxes, interfaces or print forms . Alternatively, costs may be estimated by associating cost to logic design, development, test, adjustment and documents; but, not according to the requirements. System developers estimate according to the complexity of projects, but customers expect the cost according to the number of requirements. Thus, customers could not understand the estimates provided by system developers. This paper shows a method to calculate cost share rate for each requirement in order to evaluate requirements accurately with mutual understanding of the developer and customers. Also cost share rate indicates importance of each requirement.Cost share rate is defined as the percentage of total cost assigned to each requirement (see Figure 3). Although in this research cost rate(%) for each requirements are subjective figure from the two staffs that worked on this system development project (see cost share rate in the Table 5, Table 6). And cost share rates are gained by multiply cost rate by cost under the estimate. This cost rate and estimated cost are the value that according to the category.

7 Probability Prediction of schedule delay

This paper aim to predict risk of system development project based on requirements analysis. This paper consider one of risk of system development project is schedule delay. Thus this research suppose probability of schedule delay follows β distribution(see Equation 5), then parameters (see Table 3) are gained by curve fitting sample data into β distribution. Sample data in Table 2 are surveyed in past system development project; Rdf system for tool tracking in machine factory. In this analyzing process x is probability parameter that indicates start day for

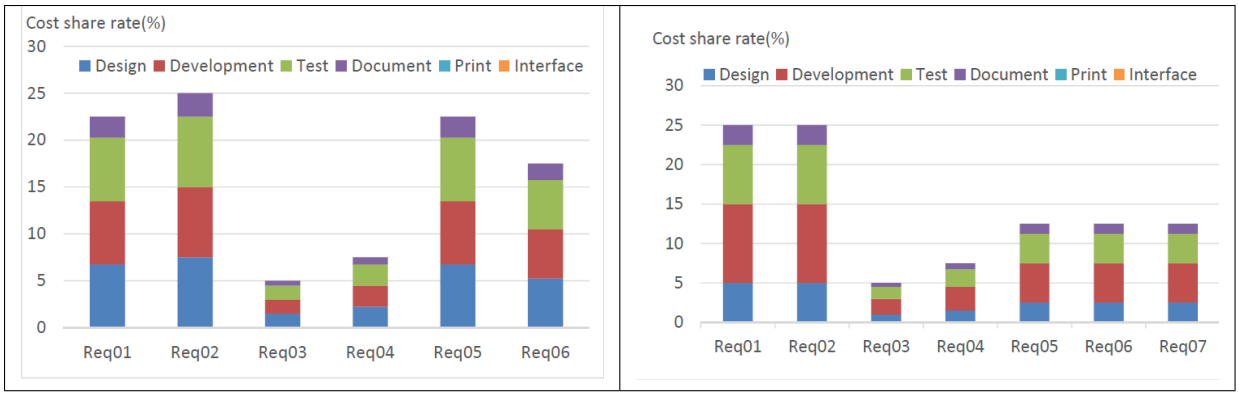


Figure 3 Left panel: Extract keywords from requirement version one and Relation with typical category. Right panel: Extract keywords from requirement version four and Relation with typical category.

Table 1 Overlap keywords from requirement version one and Relation with typical category

	Req01	Req02	Req03	Req04	Req05	Req06
Req01	0	3	3	2	2	2
Req02	3	0	4	3	3	2
Req03	3	4	0	7	3	3
Req04	2	3	7	0	2	2
Req05	2	3	3	2	0	3
Req06	2	2	3	2	3	0
Design	1	3	4	3	2	0
Development	3	2	5	4	3	4
Print	0	0	0	0	0	0
Check	0	0	0	0	0	0
Interface	0	0	0	0	0	0
Document	0	0	0	0	0	0
The number of over 4 relationship with requirement of each category are colored						

each task, and y is ratio for schedule delay against actual days. Table 2 shows that survey/preparation, design and programming process have risk of schedule delay. Although test and writing document process have no risk of schedule delay. Figure 4 shows β distribution curve in this case from parameters (see Table 3).

$$f(x) = c \cdot x^{\alpha\,-1}(1-x)^{\beta\,-1} \tag{5}$$

Figure 4 shows that probability of schedule delay is 0.225.

8 Risk Prediction for business system development project

Thus this research assign probability of risk; 0.225 to essential requirements that distinguished by requirements analysis. Actually Table 5 shows that requirement3 and requirement4 are about design and development, requirement6 is about development. These requirements are essential, and assigned probability of risk as 0.225. Requirement1 is about Main Design,and assigned probability of risk as 0.1, because in past three project contingency is set as 10% (see Table 4). Contingency is set for refinement: deleting,adding or refactoring. Table 4 shows occupying cost rate for each work process from surveying of past system development project; medical record system, knowledge management system. Thus risk is calculated probability of risk $\times$ cost (see equation 2), cost is analyzed by cost share rate. In estimated cost of requirement version one estimated total cost is 109 (see Table 5). And at completion actual total cost is 109.1 (see Table 6). At the result these results are equal.

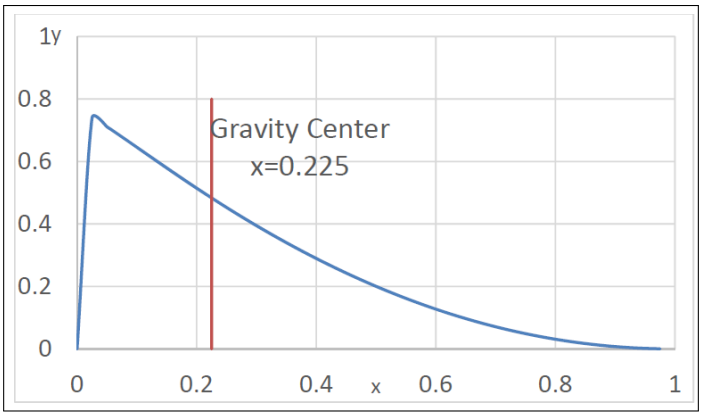


Figure 4 Beta distribution from past project

Table 2 Sample data from past project

Item	Survey/	Design			Programing			Test	Document	Sum
	Preparation	Dialog	Function	DB	Dialog	Function	DB			
Start date:x	0	0.52	0.6	0.59	0.7	0.74	0.59	0.947	0.931	
Schedule day	11	29	27	27	9	58	39	3	13	216
Actual day	35	34	36	27	9	55	62	2	7	267
Delay day	24	5	9	0	0	-3	23	-1	-6	51

Table 3 Acquire Parameter for β distribution

C	α	β
0.839835	1.020625	3.047617

Table 4 Data from past project

Project	Design	Development	Testing	Document	Refinement	Total
P1	3	4	1	1	1	10
P2	1	6	1	1	1	10
P3	1	6	1	1	1	10
Average	1.7	5.3	1	1	1	10

9 Conclusion

Meeting budget, finishing on schedule, and maintaining high quality are all important in project management. However if cost and duration are not estimated accurately, it is very difficult to meet budget and finish on fine. Additionally there are gaps between estimates by system developers and the estimates that customers are expecting, because there is a difference between system developers and customers in grouping the costs to estimate. Estimate miss by developer causes finally over cost and schedule delay.As this paper shows, usually estimate is measured by amount of source code of system or complexity of system. It is not estimated based on essentiality or risk of requirements. Misunderstanding of requirement causes misses in estimate or schedule. Then, this paper proposes a method to find essential requirements in order to estimate accurately. Additionally, distinguishing requirements that have high risk is vital in order to finish projects fine. Usually on risk management risks are evaluated according to staffs' experience. Also risks are not evaluated based requirement analysis. Evaluating risk properly is needed in order to manage project well. Thus this paper show better results by using cost share rate from requirement analysis. Also it shows potential to predict probability under the budget at completion in business system development projects. Additionally, this method helps prioritize requirements and narrow down specifications of the project.

Table 5 Estimated budget from requirement version one with risk analysis

Requirement Version one	Attribute	Cost Share Rate Ver one(%)	Schedule delay Probability	Conditional Probability	Monetary Risk (Contingency)	Estimated Cost Ver one
Requirement1	M	22.5	0	0.1	2.25	24.75
Requirement2		25	0	0	0	25
Requirement3	D/De	5	0.225	0	1.125	6.125
Requirement4	D/De	7.5	0.225	0	1.6875	9.1875
Requirement5		22.5	0	0	0	22.5
Requirement6	De	17.5	0.225	0	3.9375	21.4375
Total		100			9	109

M:Main, D:design, De:Development

Table 6 Total cost at completion

Requirement Version four	Attribute	Cost Share Rate Ver four(%)	Increased cost Rate(%)	Total Cost At Completion
Requirement1	M	25	20	27
Requirement2		25	10	27.5
Requirement3	D/De	5	0	5
Requirement4	D/De	7.5	5	7.875
Requirement5		12.5	0	22.5
Requirement6	De	12.5	10	19.25
Requirement7		12.5		0
Toatal		100		109.125

M:Main, D:design, De:Development

Prioritizing requirements and narrowing down specifications accurately help ensure it meets budget and duration targets. Additionally, this paper shows possibility to obtaining probability for risks accurately with cost share rate. However, this result was obtained by small case. Thus further research and study is needed to refine and improve this method to obtain cost share rate and risk more accurately.

10 Discussion

Customers and developers estimate costs differently, resulting in differing expectations for project cost. If estimates could be proposed to customers according to the cost of customers requirements, there would be improved understanding between system developers and customers. Because in business system development projects there are many ways to implement requirements, there are large variability in translating user's requirement into system specification. It differs greatly according to staff's skill. And it is one reason why estimate does not meet final cost. Thus proper estimate by requirement analysis is needed in order to finish business system development project fine. Final estimate is build by adding contingency to base estimate. Also correct probability for risk is needed to build proper estimate. Requirements have invisible risks. There are plus risk and minus risk. Minus risk gets prospect of profit worse, but plus risk gets prospect of profit well. But plus risk is not visible, plus risk is only in mind of staffs individually. This is one reason why estimated cost: 109 match total cost: 109 (see Table 5, Table 6). Therefore there is potential to predict risk accurately using conditional probability or Bayesian analysis [6], [7], [8] in order to visualize plus risk.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17K00354. The author would like to thank INTERLINK Inc. for data and advices. The author would like to thank TECHNO SOLUTION Inc. for data and advices.

References

[1] Mochida, S., Righter, J., Summers, J.D.: A Study of Project Cost Management Based on Requirements Analysis, *International Journal of BMFSA,* **21** (1), 2016, 21–27.

[2] Rad, P.F., Itou, K.: *Project Estimating and Cost Management*, Seisansei Syutupan, 2004.

[3] Mochida, S.: A Study on Method of Measuring Performance for Project Management, In: *20th ISPE International Conference on Concurrent Engineering,* IOS Press Ebooks, 2013, 264–273.

[4] Morkos, B., Mathieson, J., Summers, J.D. : Comparative analysis of requirements change pre-diction models: manual, linguistic, and neural net-work, *Research in Engineering Design,* **25**, Springer, 2014, 139–156.

[5] Project Management Institute: *A Guide to the Project Management Body of Knowledge,* Fourth Ed., Project Management Institute, 2008.

[6] Christian, R.: *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation,* Springer Texts in Statistics, 2007.

[7] Hoel, P.G., Asai, A., Murakami, M.: *Elementary Statistic*, Springer Texts in Statistics, 2011.

[8] Blackwell, D., Mikuni, K.: *Basic Statistic*, Kogakusha, 1972.

Bayesian Study on When to Restart Heuristic Search

Matej Mojzes [1], Jaromir Kukal [2]

Abstract. Heuristic algorithm performance measures assess the quality of a search process by statistically analyzing its performance data, typically the number of objective function evaluations before optimal or acceptable solution is found. Such criteria are not only intended to provide the verdict on which algorithm is better for what task, but also to help make the best possible use of a given algorithm on a given task. This target may be achieved by an appropriate restart strategy of the search process. In our paper we formulate axiomatic approach which also describes existing performance measures. Novelty of this paper consist in performance measure analysis via Markovian chain calculation and its direct Bayesian estimation based on Monte Carlo simulations. Practical results are demonstrated on combinatorial optimization problems and are applicable e.g. to NP-hard problems from the field of operational research.

Keywords: Heuristics, search process, performance measure, restart strategy, Bayesian analysis, operational research.

JEL classification: C44
AMS classification: 90C15

1 Introduction

The task of searching for optimal or acceptable solution depends mostly on complexity of the problem being solved. Complexity is affected particularly by shape and multi-modality of the problem objective function; by the dimension of the problem, i.e. number of input variables and last, but not least, by computational cost of objective function evaluation. There are many well-known challenging optimization problems, for example in the domain of operational research [1], that are motivating us to construct various heuristics. There are some sophisticated heuristics, e.g. Fast Simulated Annealing [2] or Cuckoo Search [3] and other, rather straightforward ones, like Steepest Descent [4].

The traditional statement of No Free Lunch Theorem (NFLT) for search / optimization suggests that over all possible objective functions, all search algorithms perform equally well and no algorithm will perform better than straightforward random search in general [5]. However, when it comes to a specific problem, it is relevant to question which algorithm is the most suitable one (or vice versa, we can compare difficulty of different problems w.r.t. specific optimization algorithm). Thus, interpretation of the NFLT motivates us to develop search algorithm performance measures and to compare different search algorithm instances on a given task [6].

While one heuristic may perform well on one type of problem in terms of reliability and speed of convergence, for an another combination of heuristic and/or task it may be more suitable to stop searching prematurely and start again [7], [8], [9], [10]. The purpose of this article is to develop the methodology and tools that will assess the performance of a given heuristic on a given task and will try to determine its optimal *restart strategy*, e.g. to decide whether it is more effective to let the heuristic evolve its best solution in one long run or to use multiple independent shorter runs. The ultimate goal and motivation for our work is to improve overall search process efficiency according to the new new search quality criterion estimation.

Keeping in mind theoretical properties of optimal restart strategies for selecting cut-off times studied by [11], we will focus on a novel and practical Bayesian estimation of a more recent performance measure [9] and we will demonstrate its potential real-world benefits. For the estimation itself we will also use heuristic run-time distribution analysis introduced mostly for descriptive-only purposes by [12], [13], [14], [15], [7]. Restart strategies are popular and often used in conjunction with some problem-specific heuristics, e.g. [16], however, our intention is to deliver a problem-independent methodology.

Last, but not least, the best way of how to choose the optimal running strategy is to have results of own experiments [8], but especially for research purposes, it is also very important to be able to make use of the results that have already been published in printed research papers. Therefore tools presented in this article are adapted for analysis of basic statistics that are usually included in relevant literature as well.

[1]FNSPE CTU, Dept. of Software Engineering, Trojanova 13, Prague, Czech Republic, mojzemat@fjfi.cvut.cz
[2]FNSPE CTU, Dept. of Software Engineering, Trojanova 13, Prague, Czech Republic, jaromir.kukal@fjfi.cvut.cz

2 The Search Process

Let $\mathbf{U}$ be a non-empty *set of states*. Let $\mathbf{G} \subset \mathbf{U}$ be a non-empty *set of goals*. Any state $\mathsf{x} \in \mathbf{G}$ is called a *solution* of the *searching task* $\langle \mathbf{U}, \mathbf{G} \rangle$. Let $N \in \mathbb{N}$ be the maximum number of searching steps. Any algorithm generating the sequence of $(\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_N) \in \mathbf{U}^N$ is called a *search process* (SP). The *number of searching steps* (time complexity of SP) is defined as $n = \min\{k \in \mathbb{N} \,|\, \mathsf{x}_k \in \mathbf{G}\}$, i.e. in the case of successful search. Should the search end with a failure we set $n = +\infty$.

A SP that is generated by deterministic algorithm will have constant complexity n. However, since this may be considered an extreme case, we will suppose that the SP is produced by a stochastic algorithm and the complexity n is a stochastic variable with the domain of $\mathbf{D} = \{1, 2, \ldots, N, +\infty\}$ and densities $p_n \geq 0, \forall n \in \mathbf{D}$ satisfying $\sum p_n = 1$. The value of p_n for $n \leq N$ may be interpreted as the probability of finding the solution in n-th step of the SP. Moreover, we may define $p_{\mathrm{succ}} = \sum_{n=1}^{N} p_n$ as the probability of success and $p_\infty = 1 - p_{\mathrm{succ}}$ as the probability of failure in a single run of SP. In the following we will be studying SP with $p_{\mathrm{succ}} > 0$ only.

Example 1. A typical example of the searching task $\langle \mathbf{U}, \mathbf{G} \rangle$ is an optimization problem, where $\mathbf{G} = \{\mathsf{x} \in \mathbf{U} \,|\, \mathrm{f}(\mathsf{x}) \leq \mathrm{f}^*\}$ with $\mathrm{f}^* = \min\{\mathrm{f}(\mathsf{x}) \,|\, \mathsf{x} \in \mathbf{U}\}$.

To further study the behaviour of a SP we may use three basic, often used, e.g. by [3], characteristics:

- p_{succ} as *reliability* of the SP,
- $\mathrm{E}\, n = p_{\mathrm{succ}}^{-1} \sum_{n=1}^{N} n\, p_n$ as *mean number of searching steps* in the case of successful search,
- $\sqrt{\mathrm{D}\, n} = p_{\mathrm{succ}}^{-1/2} \left(\sum_{n=1}^{N} (n - \mathrm{E}\, n)^2\, p_n\right)^{1/2}$ as *standard deviation of the searching step number* in the case of successful search.

To address the fundamental problem of measuring SP *time complexity*, as long as the SP has reliability of $p_{\mathrm{succ}} = 1$, a very straightforward criterion of mean number of steps $\mathrm{E}\, n$ is frequently used. On the other hand, for $0 < p_{\mathrm{succ}} < 1$ we have to adjust the value of $\mathrm{E}\, n$ due to decreased reliability of the SP. An example of adjusted time complexity evaluation is based on the Feoktistov criterion [17] $FEO = \mathrm{E}\, n / p_{\mathrm{succ}}$. Our first aim is to develop a new time complexity criterion in the case of infinite time period which will be the result of mean value calculations as well. Our second aim is to develop an alternative criterion of SP quality in the case of constrained searching time. However, to begin, we will define the Extended Search Process which will allow us to measure the time complexity.

3 Extended Search Process

If the SP is successful in the first run, then the searching task is done. Otherwise, should the process end with a failure, we continue to repeat new runs until we succeed. This trivial but practical habit can be called *Extended Search Process* (XSP). Our methodology is based on the following three axioms.

Axiom 1. The only one possibility of how to guarantee $p_{\mathrm{succ}} = 1$ is by substituting SP with unconstrained XSP.

Axiom 2. If $p_{\mathrm{succ}} = 1$ the mean value of number of steps is the only one acceptable criterion of SP quality.

Axiom 3. If the number of XSP steps is constrained by $n \leq M$, the only one acceptable criterion of SP quality is the reliability of XSP for $n = M$.

The first axiom "legalizes" the necessity of SP repeating to achieve theoretical completeness [8] and also due to lack of any other reasonable alternative. Additionally, under the first axiom the resulting search algorithm may be regarded as a Las Vegas algorithm [18], which is often used to analyze heuristic performance [11], [8]. However, the other axioms are subjective. In the case of the second axiom, one could use a family of other characteristics (median, quantiles, mean logarithm, etc.), but their theoretical investigation would be more complicated and hardly applicable for analysis of other authors results (which are often published using the mean number of steps and reliability). The third axiom corresponds to the situation when we are prepared to spend the time of M steps of XSP at most. In this case motivation for raising the M value is in potential improvement of XSP reliability as the criterion of XSP quality.

Therefore, the XSP will be analysed for both unconstrained and constrained number of steps. At first, in the following, we will suppose Axiom 1 and Axiom 2 do hold and thus the XSP is unconstrained.

Let $k \in \mathbb{N}$ be the index of an SP run. Let $j = 1, 2, \ldots, N$ be the searching step index inside the individual SP run. The distinct runs of SP are supposed to be independent and, therefore, the XSP successfully terminates in the $n-\mathrm{th}$ step with probability $p_n^* = p_{N(k-1)+j}^* = (1 - p_{\mathrm{succ}})^{k-1} p_j$, where p_j is the probability of finding the solution in j-th step of the SP. It is easy to verify that XSP is reliable and thus, the mean value of searching step number for successful search in XSP is the only one criterion of the XSP time complexity and, consequently, also

of the SP time complexity. Furthermore, we can calculate

$$\mathrm{E}\, n^* = \frac{N\, p_{\mathrm{succ}}\,(1 - p_{\mathrm{succ}})}{p_{\mathrm{succ}}^2} + \frac{p_{\mathrm{succ}}\,\mathrm{E}\, n}{p_{\mathrm{succ}}}\,. \tag{1}$$

Hence, the relationship between time complexities of the XSP and the original SP is

$$\mathrm{E}\, n^* = \mathrm{E}\, n + N \cdot \frac{1 - p_{\mathrm{succ}}}{p_{\mathrm{succ}}}\,. \tag{2}$$

Finally, this formula can be used directly to build up the XSP quality criterion and, as has been noted, of SP time complexity

$$Q_\infty = \mathrm{E}\, n + N \cdot (p_{\mathrm{succ}}^{-1} - 1)\,. \tag{3}$$

It is obvious that the new criterion is quite similar to the Feoktistov's one. Let us denote exact relation between the two. Starting from the inequality of $n \leq N$ we obtain $\mathrm{E}\, n \leq N$ and then $Q_\infty \geq \mathrm{E}\, n + \mathrm{E}\, n \cdot (p_{\mathrm{succ}}^{-1} - 1) = \mathrm{E}\, n / p_{\mathrm{succ}} = FEO$. We can conclude that, in the case of unconstrained XSP, Feoktistov's criterion is the lower bound of the Q_∞ criterion of SP time complexity.

We can also study the SP via failure probability $s_j = 1 - \sum_{k=1}^{j} p_k$ for $j = 0, 1, \ldots, N$. Therefore, we obtain $s_j^* = s_{N(k-1)+j}^* = s_N^{k-1} s_j$ where $1 \leq j \leq N$, s_N^{k-1} is failure probability after $k-1$ complete independent runs of SP, s_j is failure probability after j-th step of SP during k-th run, and finally $Q_\infty = 1 + \frac{\sum_{j=1}^{N} s_j}{1 - s_N}$.

In the second case, we suppose Axiom 1 and Axiom 3 hold. Also, we suppose that a single run of SP is constrained by N, while, on the other hand, M is the constraint for entire XSP. Thus, the XSP is called constrained and its reliability is $\sum_{n=1}^{M} p_n^* = 1 - s_M^*$ which has to be maximal from the point of view of real application. An equivalent aim would be to minimize the failure probability s_M^* of XSP. According to the relationship between SP and XSP we directly have the quality criterion in the constrained case, Q_M, defined as $Q_M = s_N^{k-1} s_j$, where $k = 1 + \lfloor \frac{M-1}{N} \rfloor, j = M - (k-1)N$.

In the case of trimmed SP of length $n \leq N$ XSP quality analysis for $M \in \mathbb{N} \cup \{+\infty\}$ offers the main yield. Therefore Q_M is a function of $n \in \{1, 2, \ldots, N\}$ and we can easily find $Q_{\mathrm{opt}} = \min\limits_{1 \leq n \leq N} Q_M(n)$ and $n_{\mathrm{opt}} = \min\{n \,|\, Q_M(n) = Q_{\mathrm{opt}}\}$.

4 Direct estimation of Q_∞

Theoretical analysis of heuristics, if possible at all, is often very limited for practical applications. Instead, heuristics are often studied via empirical methodology based on performing M independent runs of SP and recording the time required to find a solution. Clearly, the more runs are performed, the better will the estimate of original SP [8].

The M independent runs of SP which yield time complexities $n_1, \ldots, n_M \in \{1, \ldots, N, +\infty\}$ of individual runs. Denoting $M^* = \mathrm{card}\{k \,|\, n_k < +\infty\}$ as the number of successful runs, $\mathrm{E}^* = \frac{1}{M^*} \sum_{n_k < +\infty} n_k$ for $M^* \geq 1$ and $\mathrm{D}^* = \frac{1}{M^* - 1} \sum_{n_k < +\infty} (n_k - E^*)^2$ for $M^* \geq 2$, we have E^*, D^* as unbiased estimates of $\mathrm{E}\, n$ and $\mathrm{D}\, n$, the basic descriptive statistics of the run-time required for obtaining a given solution quality.

The novel estimation of Q_∞ makes use of Bayesian approach [19] and properties of Dirichlet distribution [20]. According to (3), unbiased estimation of Q_∞ is based on the term $p_{\mathrm{succ}}^{-1} - 1$ and Bayesian estimation of its mean value and variation. Beginning with prior uniform distribution of $p_{\mathrm{succ}} \in (0, 1)$ we obtain Dirichlet posterior distribution with density $\mathrm{Di}(p_{\mathrm{succ}}) = \frac{(M+1)!}{M^*!(M-M^*)!} \cdot p_{\mathrm{succ}}^{M^*} \cdot (1 - p_{\mathrm{succ}})^{M - M^*}$, and then we obtain Bayesian estimates of mean

$$\mathrm{E}(p_{\mathrm{succ}}^{-1} - 1) = \frac{M - M^* + 1}{M^*} \tag{4}$$

for $M^* \geq 1$ and the variance

$$\mathrm{D}(p_{\mathrm{succ}}^{-1} - 1) = \frac{(M+1)(M - M^* + 1)}{(M^*)^2 (M^* - 1)} \tag{5}$$

for $M^* \geq 2$. Therefore, the unbiased estimate of Q_∞ is

$$Q_\infty^* = E^* + N \cdot \frac{M - M^* + 1}{M^*} \tag{6}$$

for $M^* \geq 1$ with standard deviation

$$s_\infty^* = \sqrt{\frac{\mathrm{D}^*}{\mathrm{M}^*} + \frac{N^2}{M^*} \cdot \frac{(M+1)(M - M^* + 1)}{(M^*)^2 (M^* - 1)}} \tag{7}$$

for $M^* \geq 2$. Resulting formulas (6, 7) are the novel Bayesian estimates, which enable more sophisticated run-time analysis. Main application of (6) is in the experimental optimization of the maximum SP length, or cut-off time [8], N.

Example 2. Let $d, a \in \mathbb{N}$ be the dimension and range of optimization and $A \in \mathbb{R}^{d\times d}$ be a regular matrix. Then $\mathbf{U} = \{\mathbf{x} \in \mathbb{R}^d \mid ||\mathbf{x}||_\infty \leq a\}$, and we will minimize objective function $\mathrm{f}(\mathbf{x}) = ||A\mathbf{x}||_1$ on the domain $\mathbf{U}$. The optimal value is $\mathrm{f}^* = \mathrm{f}(\mathbf{0}) = 0$ and the $\mathbf{G} = \{\mathbf{0}\}$. Two states $\mathbf{x}, \mathbf{y} \in \mathbf{U}$ are defined as neighbors when $||\mathbf{x} - \mathbf{y}||_1 \leq 1$. In this case we can apply simple optimization heuristic called *Steepest Descent* [4] as $\mathbf{x}_1 = \texttt{init}(\mathbf{U})$ and $\mathbf{x}_{k+1} = \mathbf{y}_k$ for $\mathrm{f}(\mathbf{y}_k) < \mathrm{f}(\mathbf{x}_k)$ or $\mathbf{x}_k$ otherwise, where $\mathbf{y}_k$ is the best neighbor of $\mathbf{x}_k$.

The resulting SP can be studied as a Markov chain [21] with a finite number of states if the $p_1, p_2, \ldots, p_N$ probabilities can be calculated for a given N. A numerical study was performed for $d = 6$, $a = 1$, $A = \mathsf{H}^{-1}$ which is an inversion of the Hilbert matrix [22].

For this task we have performed $M = 1000$ independent runs of SP for various length constraints N. The results are collected in Table 1. The theoretical optimal termination point is $N_{\mathrm{opt}} = 23$ and the experimental estimate $N^*_{\mathrm{opt}} = 24$ which is near the theoretical value. We can conclude that, although being quite limited, this example proves practical relevance of Q_∞ estimation.

N	M^*	E^*	D^*	Q^*_∞	s^*_∞	Q_∞
20	78	10.667	28.693	247.333	3.238	259.621
21	83	11.289	33.062	243.554	3.007	258.419
22	88	11.898	37.380	240.148	2.808	257.751
23	90	12.144	39.249	244.956	2.806	**257.560**
24	98	13.112	46.657	**234.255**	2.486	257.796
25	101	13.465	49.371	236.238	2.440	258.415

Table 1 Theoretical value of Q_∞ and its estimation Q^*_∞ – optimal values (in bold) are almost the same

Example 3. Clerc's Zebra3 problem is a non-trivial binary optimization problem and part of discrete optimization benchmark problems [23]. For this task we have chosen to use the physically-motivated Fast Simulated Annealing (FSA) [2] with reputable efficiency in the case of integer optimization tasks. Practical effectiveness of this heuristic depends mostly on selection of appropriate temperature and cooling strategy. Thus, we set up an instance with annealing temperature $T = 0.283$. Furthermore, we set $N = 10000$ and ran the instance 1000 times without any restarting.

After this experiment we analyzed time complexity of the task in order to estimate corresponding N^*_{opt}. Analyzed instance turned out to be very slow and unreliable, with $p_{\mathrm{succ,A}} = 0.29$ only – i.e. characteristics, that are comparable with pure random search. Nevertheless, the suggested cut-off time was $N^*_{\mathrm{opt,A}} = 16$, i.e. almost immediately after start. In the next step we repeated the search task the same 1000 times with $N = 10000$, but the heuristic sub-run was allowed to run only for N^*_{opt} steps at most. The resulting SP is considerably faster and substantially more reliable – improvement is clearly visible on Figure 1. All numerical results for this example are collected in Table 2.

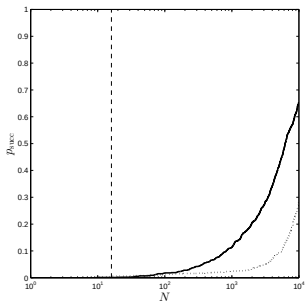


Figure 1 Reliability based on N for FSA instance A, dashed line describes the run without restarting and solid line the run restarted using $N^*_{\mathrm{opt}} = 16$, which is also marked by the vertical line. It is clear that by restarting the search in its first steps the overall SP speed and reliability are significantly improved.

Based on this example we may conclude that a sophisticated heuristic, even if it is apparently not performing better than random search on a given task, may be improved noticeably just by its restarting at the right time.

N^*_{opt}	$\mathrm{E}\, n$	$\sqrt{\mathrm{D}\, n}$	p_{succ}	Q^*_∞	s^*_∞
N/A	6091	2875	29	31091	1760
16	4039	2810	65	9369	371

Table 2 Basic characteristics and complexity estimates of FSA on Clerc's Zebra3. Where N^*_{opt} is not available, it means that search was limited only by N, i.e. not restarted and used to estimate N^*_{opt} for the run with restarts.

Complexity estimate $Q^*_\infty \pm s^*_\infty$ can also be used for tuning the other parameters (N is only one of them) of a given heuristic on a given task. Another application of estimate (6) is in the comparison of heuristics or their settings on a specific task.

Many authors publish their results introducing M, N and then they calculate only E^* and reliability $r = M^*/M$. Having two heuristics, say: A and B, we can directly calculate $Q_{\infty,\mathrm{A}}$, $Q_{\infty,\mathrm{B}}$ and recognize that algorithm A is better than algorithm B only if $Q_{\infty,\mathrm{A}} < Q_{\infty,\mathrm{B}}$. However, this naive approach may be used only in the case of missing information on the variance D^*.

Example 4. A good example is the comparison of two optimization heuristics based on differential evolution (DE): modified composite DE (A) and self-adapting DE (B) on the task consisting in minimization of Schwefel function [24]. In here, $M = 100, N = 200000$. In the case of the first heuristic we have $E^*_{\mathrm{A}} = 20207, r_{\mathrm{A}} = 1.00$ and then $M^* = 100$ and $Q^*_{\infty,\mathrm{A}} = 22207$. On the other hand, the latter heuristic performance may be described by $E^*_{\mathrm{B}} = 8557, r_{\mathrm{A}} = 0.98$ and thus $Q^*_{\infty,\mathrm{B}} = 14679.45$. Since there is no additional information on the variances, we may declare that in this case self-adapting DE is better than modified composite DE.

Such approach can be used not only to compare two distinct heuristics, but also to compare multiple heuristics on a given task without any additional effort.

On the other hand, in many other papers authors publish the estimates of standard deviation $s = \sqrt{D^*}$. Therefore we can calculate also $s^*_{\infty,\mathrm{A}}$ and $s^*_{\infty,\mathrm{B}}$ according to (7), and then perform statistical testing using standard z-score technique as

$$z = \frac{|Q^*_{\infty,\mathrm{A}} - Q^*_{\infty,\mathrm{B}}|}{\sqrt{(s^*_{\infty,\mathrm{A}})^2 + (s^*_{\infty,\mathrm{B}})^2}} \tag{8}$$

$$p_{\mathrm{value}} = 2 - 2\Phi(z)\ , \tag{9}$$

where Φ stands for cumulative distribution function of Gaussian distribution.

Example 5. Griewank's function is used in [3] to compare two optimization heuristics: Cuckoo Search (CS) – A and Particle Swarm Optimization (PSO) – B. Having $M = 100, N = 100000$, $E^*_{\mathrm{A}} = 10912, r_{\mathrm{A}} = 1.00, s_{\mathrm{A}} = 4050$ and $E^*_{\mathrm{B}} = 55970, r_{\mathrm{B}} = 0.92, s_{\mathrm{B}} = 4223$, we can calculate $Q^*_{\infty,\mathrm{A}} = 11912, s^*_{\infty,\mathrm{A}} = 417.41$, $Q^*_{\infty,\mathrm{B}} = 65752.61, s^*_{\infty,\mathrm{B}} = 567.56$, and finally $z = 76.42, p_{\mathrm{value}} = 2.75 \times 10^{-1271}$. Therefore CS is significantly better than PSO in the case of minimization of Griewank's function and level $\alpha = 0.05$.

Multiple heuristics comparison is also possible – having H heuristics we have to perform $H(H-1)/2$ pair comparisons. Beginning with N_j, M_j, E^*_j, r_j and s_j, we calculate M^*_j, D^*_j and then $Q^*_{\infty,j}, s^*_{\infty,j}$ for $j = 1, \ldots, H$. Multiple pair comparison generates $z_{i,j}$ and $p_{\mathrm{value,i,j}}$ according to (8), (9) for $1 \le i < j \le H$. The methodology of *False Discovery Rate* (FDR) [25] enables us to recognize significant differences in heuristics quality measure Q^*_∞ [26].

5 Conclusion

In this paper, we had established axioms and derived heuristic performance criteria in a way that enabled us to elaborate their novel Bayesian estimation and finally examine the performance of a given heuristic algorithm on a given task. We had introduced and demonstrated how this may be accomplished by both using the data from own experiments or by extracting the data from existing papers. If applying the latter and probably more practical option, we recommend to use the Q_∞ measure, and otherwise Q_M with some reasonable ceiling M.

To sum up, knowing E^* and r, we may still compare the respective Q_∞ measures, although not in the statistical sense. Moreover, knowing E^*, r, and s, we can statistically test values of Q_∞. And what is more, with such data we are able to estimate N_{opt} minimizing Q_∞.

The cases above are perhaps mainly useful for termination time analysis of someone else's heuristic. On the other hand, having full experimental data, we are able to determine N_{opt} and optimal running strategy in the most straightforward way.

Acknowledgements

This paper was created under the support of grant SGS17/196/OHK4/3T/14 Czech Technical University in Prague.

References

[1] Cook, W.J. et al: Combinatorial Optimization. Wiley (1997)

[2] Černý, V.: Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. J. Optimization Theory and Applications, Kluwer Academic Publishers-Plenum Publishers, 45, 41–51 (1985)

[3] Yang, X.-S., Deb, S.: Cuckoo search via Lévy flights. In: Proc. of World Congress on Nature & Biologically Inspired Computing, pp. 210–214, IEEE Publications (2009)

[4] Snyman, J.A.: Practical Mathematical Optimization: An introduction to basic optimization theory and classical and new gradient-based algorithms. Springer (2005)

[5] Wolpert, D.H., Macready, W.G.: No Free Lunch Theorems for Optimization. IEEE Transactions on Evolutionary Computation 1, 67–82 (1997)

[6] Gendreau, M., Potvin, J.Y.: Handbook of Metaheuristics, International Series in Operations Research & Management Science. Springer (2010)

[7] Hoos, H.H., Stützle, T.: Evaluating Las Vegas Algorithms - Pitfalls and Remedies. In: Proc of Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI-98), pp. 238-–245 (1998)

[8] Hoos, H.H., Stützle, T.: Stochastic Local Search: Foundations and Applications. Morgan Kaufmann Publishers (2005)

[9] Auger, A., Hansen, N.: Performance evaluation of an advanced local search evolutionary algorithm, In: Proc of IEEE Congress on Evolutionary Computation (CEC 2005), pp. 1777 -– 1784 (2005)

[10] Hansen, N., Auger, A., Finck, S., Ros, R.: Real-parameter black-box optimization benchmarking 2010: Experimental setup, (2010)

[11] Luby, M., Sinclair, A., Zuckerman, D.: Optimal speedup of Las Vegas algorithms, Information Processing Letters, 47, 173–180 (1993)

[12] Taillard, É. D.: Robust taboo search for the quadratic assignment problem. Parallel Computing, 17, 443-–455 (1991)

[13] Battiti, R., Tecchiolli, G.: Parallel biased search for combinatorial optimization: Genetic algorithms and TABU. Microprocessors and Microsystems, 16, 351-–367 (1992)

[14] Taillard, É. D.: Parallel tabu search techniques for the job shop scheduling problem. ORSA Journal on Computing, 6, 108-–117 (1994)

[15] Ten Eikelder, H., M., M., Verhoeven, M., G., A., Vossen, T., V., M., Aarts, E., H., L.: A probabilistic analysis of local search. In: Meta-Heuristics, 605–618, Springer US (1996)

[16] Guo, Y., Zhang, B., Zhang, C.: A heuristic restart strategy to speed up the solving of satisfiability problem. In Computational Intelligence and Design (ISCID), 2012 Fifth International Symposium on, 2, pp. 423–426 (2012)

[17] Feoktistov, V.: Differential Evolution: In Search of Solutions. Springer (2006)

[18] Babai, L.: Monte-Carlo algorithms in graph isomorphism testing, Université tde Montréal Technical Report, DMS, 79-10 (1979)

[19] Howson, C., Urbach P.: Scientific Reasoning: the Bayesian Approach. Open Court (2005)

[20] Balakrish, V.B.: Dirichlet Distribution. In: Balakrishnan, N., Nevzorov, V. B.: A Primer on Statistical Distributions, p. 274, John Wiley & Sons (2005)

[21] Meyn, S., Tweedie, R.L.: Markov Chains and Stochastic Stability. Cambridge University Press (2009)

[22] Hilbert, D.: Ein Beitrag zur Theorie des Legendre'schen Polynoms. Acta Mathematica, 18, 155-–159 (1894)

[23] Hanning, C., Yunlong, Z., Kunyuan, H., Xiaoxian, H.: Hierarchical Swarm Model: A New Approach to Optimization, Discrete Dynamics in Nature and Society, Vol.2010 (2010).

[24] Tvrdík, J., Poláková, R., Bujok, P.: A Comparison of Adaptive Differential Evolution Variants for Single-Objective Optimization. In: Proc. of Mendel 2012 Soft Computing Conference, pp. 132–137, Brno Univ Technology Press (2012)

[25] Benjamini, Y., Hochberg, Y.: Controlling the False Discovery Rate: A practical and Powerful Approach to Multiple Testing. J. R. Statistical Society 57, 289-–300 (1995)

[26] Mojzeš, M., Kukal, J., Tran, V.Q., Jablonský, J.: Performance Comparison of Heuristic Algorithms via Multicriteria Decision Analysis. In: Proc. of Mendel 2011 Soft Computing Conference, pp. 244–251, Brno Univ Technology Press (2011)

Robustness of interval Monge matrices in max-min algebra

M. Molnárová[1]

Abstract. Robustness of interval Monge matrices over max-min algebra is studied. The max-min algebra is an extremal algebra with operations maximum and minimum. An interval matrix is in fact a set of matrices between a lower bound matrix and an upper bound matrix.
Special class of matrices, namely "gamma" matrices was introduced. It was shown that a "gamma" matrix has the Monge property. Equivalent conditions for robustness of a "gamma" matrix with fixed data was proved. Polynomial algorithm for verifying the type of a given matrix and checking the robustness was introduced. Equivalent conditions for possible robustness and universal robustness as well, of an interval "gamma" matrix were proved. Polynomial algorithms for verifying the possible and universal robustness were introduced.

Keywords: (max, min) algebra, robustness, Monge matrix, interval matrix

JEL classification: C02
AMS classification: 08A72, 90B35, 90C47

1 Introduction

To model applications in many divers areas as discrete dynamic systems (DDS), graph theory, knowledge engineering or description of technical devices are the so-called extremal algebras predestinated [3]. In individual cases are addition and multiplication in linear algebra replaced by double of operations, under which at least one has the property not creating a new element. The most frequented extremal algebras are max-plus algebra (with operations maximum and addition) and max-min algebra (with operations maximum and minimum). The last above mentioned is called also fuzzy algebra. The Monge matrices and their applications were studied in [1], [4]. Robust matrices over fuzzy algebra were investigated in [13]. Sufficient and necessary conditions for robustness of Monge fuzzy matrices were proved in [7]. Robustness of interval fuzzy matrices was studied in [11], [10], [8], [9].

2 Preliminaries

The fuzzy algebra $\mathcal{B}$ is a triple $(B, \oplus, \otimes)$, where $(B, \leq)$ is a bounded linearly ordered set with binary operations *maximum* and *minimum*, denoted by $\oplus$, $\otimes$. The least element in B will be denoted by O, the greatest one by I. By $\mathbb{N}$ we denote the set of all natural numbers. The greatest common divisor of a set $S \subseteq \mathbb{N}$ is denoted by $\gcd S$, the least common multiple of the set S is denoted by $\operatorname{lcm} S$. For a given natural $n \in \mathbb{N}$, we use the notation N for the set of all smaller or equal positive natural numbers, i.e., $N = \{1, 2, \ldots, n\}$.

For any $m, n \in \mathbb{N}$, $B(m, n)$ denotes the set of all matrices of type $m \times n$ and $B(n)$ the set of all n-dimensional column vectors over $\mathcal{B}$. The matrix operations over $\mathcal{B}$ are defined formally in the same manner (with respect to $\oplus$, $\otimes$) as matrix operations over any field. The rth power of a matrix $A \in B(n, n)$ is denoted by A^r, with elements a_{ij}^r. For $A, C \in B(n, n)$ we write $A \leq C$ if $a_{ij} \leq c_{ij}$ holds for all $i, j \in N$.

A *digraph* is a pair $G = (V, E)$, where V, the so-called vertex set, is a finite set, and E, the so-called edge set, is a subset of $V \times V$. A digraph $G' = (V', E')$ is a subdigraph of the digraph G (for brevity $G' \subseteq G$), if $V' \subseteq V$ and $E' \subseteq E$. A path in the digraph $G = (V, E)$ is a sequence of vertices $p = (i_1, \ldots, i_{k+1})$ such that $(i_j, i_{j+1}) \in E$ for $j = 1, \ldots, k$. The number k is the length of the path p and is denoted by $\ell(p)$. If $i_1 = i_{k+1}$, then p is called a cycle. For a given matrix $A \in B(n, n)$ the symbol $G(A) = (N, E)$ stands for the complete, edge–weighted digraph associated with A, i.e. the vertex set of $G(A)$ is N, and the capacity of any edge $(i, j) \in E$ is a_{ij}. In addition, for given $h \in B$, the *threshold digraph* $G(A, h)$ is the digraph $G = (N, E')$ with the vertex set N and the edge set $E' = \{(i, j);\ i, j \in N,\ a_{ij} \geq h\}$.

By a *strongly connected component* of a digraph $G(A, h) = (N, E)$ we mean a subdigraph $\mathcal{K} = (N_{\mathcal{K}}, E_{\mathcal{K}})$ generated by a non-empty subset $N_{\mathcal{K}} \subseteq N$ such that any two distinct vertices $i, j \in N_{\mathcal{K}}$ are contained in a common cycle, $E_{\mathcal{K}} = E \cap (N_{\mathcal{K}} \times N_{\mathcal{K}})$ and $N_{\mathcal{K}}$ is the maximal subset with this property. A strongly connected component $\mathcal{K}$ of a digraph is called non-trivial, if there is a cycle of positive length in $\mathcal{K}$. For any non-trivial strongly connected component $\mathcal{K}$ is the *period* of $\mathcal{K}$ defined as $\operatorname{per} \mathcal{K} = \gcd \{\, \ell(c);\ c \text{ is a cycle in } \mathcal{K},\ \ell(c) > 0 \,\}$. If $\mathcal{K}$ is trivial, then $\operatorname{per} \mathcal{K} = 1$. By $\mathrm{SCC}^{\star}(G)$ we denote the set of all non-trivial strongly connected components of G.

[1]Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, Monika.Molnarova@tuke.sk

Let $A \in B(n,n)$ and $x \in B(n)$. The sequence $O(A,x) = \{x^{(0)}, x^{(1)}, x^{(2)}, \dots, x^{(n)}, \dots\}$ is the orbit of $x = x^{(0)}$ generated by A, where $x^{(r)} = A^r \otimes x^{(0)}$ for each $r \in \mathbb{N}$.
For a given matrix $A \in B(n,n)$, the number $\lambda \in B$ and the n–tuple $x \in B(n)$ are the so-called *eigenvalue* of A and *eigenvector* of A, respectively, if they are the solution of the *eigenproblem* for matrix A, i.e. they satisfy the equation $A \otimes x = \lambda \otimes x$. The corresponding *eigenspace* $V(A,\lambda)$ is defined as the set of all eigenvectors of A with associated eigenvalue λ, i.e. $V(A,\lambda) = \{x \in B(n);\ A \otimes x = \lambda \otimes x\}$.
Let $\lambda \in B$. A matrix $A \in B(n,n)$ is *ultimately λ–periodic* if there are natural numbers p and R such that the following holds: $A^{k+p} = \lambda \otimes A^k$ for all $k \geq R$. The smallest natural number p with above property is called the period of A, denoted by $\mathrm{per}(A,\lambda)$. In case $\lambda = I$ we denote $\mathrm{per}(A,I)$ by abbreviation $\mathrm{per}\, A$.

Definition 1. Let $A = (a_{ij}) \in B(n,n)$, $\lambda \in B$. Let $T(A,\lambda) = \{x \in B(n);\ O(A,x) \cap V(A,\lambda) \neq \emptyset\}$. A is called *λ–robust* if $T(A,\lambda) = B(n)$. A λ–robust matrix with $\lambda = I$ is called *a robust matrix*.

In our considerations we will use the following result (adapted for $\lambda = I$) proved in [13] to study robustness of a matrix.

Lemma 1. [13] Let $A = (a_{ij}) \in B(n,n)$. Then A is robust if and only if $\mathrm{per}\, A = 1$.

3 Robustness of Monge matrix with fixed data

In this section we recall the equivalent conditions for a Monge matrix to be robust. Afterwards we introduce a special type of a Monge matrix and prove necessary and sufficient conditions for this type of matrix to be robust. Moreover, the corresponding computational complexity of the verifying the robustness is shown to be $O(n^2)$ in this case.

Definition 2. We say, that a matrix $A = (a_{ij}) \in B(m,n)$ is a convex Monge matrix (concave Monge matrix) if and only if

$$a_{ij} \otimes a_{kl} \leq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l$$
$$\big(\, a_{ij} \otimes a_{kl} \geq a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l \,\big).$$

In this paper, we assume that the considered matrices are convex.

Equivalent conditions for robustness of Monge matrices in general case of max-min algebra were proved.

Theorem 1. [7] *Let $A \in B(n,n)$ be a Monge matrix. Then A is robust if and only if for each $h \in H$ the digraph $G(A,h)$ contains at most one non-trivial strongly connected component and this has a loop.*

Now, we introduce the matrix of type Γ and show that it has the Monge property. This fact allows us to prove equivalent conditions for this class of matrices to be robust using Theorem 1 and on the other hand to decrease the computational complexity of checking the robustness.

Definition 3. Let $A \in B(n,n)$. Let $\gamma \in N$. We say that the matrix A is
a) of type $\lceil\Gamma$, if

$$a_{ij} = \begin{cases} \text{arbitrary} & \text{if} \quad i = \gamma,\ j \geq i \\ \text{arbitrary} & \text{if} \quad j = \gamma,\ i \geq j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

b) of type $\Gamma\rfloor$, if

$$a_{ij} = \begin{cases} \text{arbitrary} & \text{if} \quad i = \gamma,\ j \leq i \\ \text{arbitrary} & \text{if} \quad j = \gamma,\ i \leq j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The matrix A is of type Γ, if it is of type $\lceil\Gamma$ or $\Gamma\rfloor$.

Remark 1. It is obvious that if the matrix A is of type Γ at most one diagonal element $a_{kk} > 0$ for $k \in N$. There are cases that a matrix A is of both types at the same time. If $a_{ii} = 0$ for all $i \in N$ there are three possibilities of ambiguity. First, if A consists of 0 entirely, then γ can be arbitrary. Second, if there is exactly one element $a_{ij} > 0$, for $i < j$, then A is of type $\lceil\Gamma$ with $\gamma = i$ and of type $\Gamma\rfloor$ with $\gamma = j$. Third, if there is exactly one element $a_{ij} > 0$, for $i > j$, then A is of type $\Gamma\rfloor$ with $\gamma = i$ and of type $\lceil\Gamma$ with $\gamma = j$. If there is a diagonal element $a_{kk} > 0$ for $k \in N$ and $a_{ij} = 0$ for all $i,\ j \in N$ except $i = j = k$, then A is of both types with $\gamma = k$. Without lost of generality, we will assume that A is of type $\lceil\Gamma$ with corresponding γ (in first case we set $\gamma = 1$) in all above cases.

Remark 2. For arbitrary $h \in H = \{a_{ij};\, i,j \in N\}$ the threshold digraph $G(A,h)$ of a matrix A of type Γ with $\gamma \in N$ contains only arcs crossing node γ. Consequently it has at most one non-trivial strongly connected component and this can contain elementary cycles of length two and of length one (a loop on node γ) only.

Lemma 2. Let $A \in B(n,n)$ be of type Γ. Then A is a Monge matrix.

Proof. Let $i,\ j,\ k,l \in N$ such that $i < k$ and $j < l$ be arbitrary but fixed. Then by Definition 3 at least one element of a_{ij} and a_{kl} equals 0. It implies that $a_{ij} \otimes a_{kl} = 0 \leq a_{il} \otimes a_{kj}$ and the assertion follows by Definition 2. □

Theorem 2. *Let $A \in B(n,n)$ be a matrix of type Γ. Then A is robust if and only if for each $h \in H$ holds: If there is $j \in N$ such that $a_{\gamma j} \otimes a_{j\gamma} \geq h$ then $a_{\gamma\gamma} \geq h$.*

Proof. Let us assume that A is robust. Let $h \in H$ be arbitrary but fixed. Let $a_{\gamma j} \otimes a_{j\gamma} \geq h$ for some $j \in N$ and let $a_{\gamma\gamma} < h$. Thus by Remark 2 the only non-trivial strongly connected component of the digraph $G(A,h)$ contains at least one cycle of length two and no loop. This is by Lemma 2 and Theorem 1 a contradiction with robustness of the considered matrix A.
The converse implication is a consequence of Remark 2 and Theorem 1. □

Theorem 3. *There is an algorithm with computational complexity $O(n^2)$ for verifying the type of a given matrix $A \in B(n,n)$ and checking the robustness.*

Proof. The verifying that the matrix is of type Γ takes $O(n^2)$ time. Since H contains at most $2n$ different elements and to verify the inequalities takes $O(n)$ time for each $h \in H$ the entire computational complexity is $O(n^2)$. □

4 Robustness of Monge matrix with inexact data

In this section we shall investigate robustness of matrices which elements are inexact, namely they are given by intervals, i.e. we will consider interval fuzzy matrices. Similarly to [2], [5], we define an interval matrix $\mathbf{A}$.

Definition 4. Let $\underline{A}, \overline{A} \in B(n,n)$, $\underline{A} \leq \overline{A}$. An interval matrix $\mathbf{A}$ with bounds $\underline{A}$ and $\overline{A}$ is defined as follows

$$\mathbf{A} = [\underline{A}, \overline{A}] = \left\{ A \in B(n,n);\ \underline{A} \leq A \leq \overline{A} \right\}.$$

An interval matrix $\mathbf{A}$ is in fact a set of matrices. Thus the question is not whether the matrix $\mathbf{A}$ is robust, but whether at least one of the matrices in $\mathbf{A}$ is robust or whether all of the matrices in $\mathbf{A}$ are robust.

Definition 5. An interval matrix $\mathbf{A}$ is called
- *possibly robust* if there exists a matrix $A \in \mathbf{A}$ such that A is robust,
- *universally robust* if each matrix $A \in \mathbf{A}$ is robust.

Definition 6. An interval matrix $\mathbf{A}^M$ for $\mathbf{A} = [\underline{A}, \overline{A}]$ is called interval Monge, if $\underline{A},\ \overline{A} \in B(n,n)$ are Monge matrices and $\mathbf{A}^M = \{A \in \mathbf{A};\ A \text{ is Monge}\}$.

Since $\underline{A},\ \overline{A} \in \mathbf{A}^M$, the set $\mathbf{A}^M$ is non-empty.

Definition 7. An interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ is called an interval matrix of type $\lceil\Gamma$ (of type $\Gamma\rfloor$) if $\underline{A}$ and $\overline{A}$ are both of type $\lceil\Gamma$ (of type $\Gamma\rfloor$) with same $\gamma \in N$. An interval matrix is of type Γ if it is of type $\lceil\Gamma$ or $\Gamma\rfloor$.

Lemma 3. An interval matrix $\mathbf{A}$ of type Γ is an interval Monge matrix.

Proof. We shall prove that $\mathbf{A} = \mathbf{A}^M$. By Definition 7 are both $\underline{A}$ and $\overline{A}$ of type $\lceil\Gamma$ (of type $\Gamma\rfloor$), hence every matrix $A \in \mathbf{A}$ is of type $\lceil\Gamma$ (of type $\Gamma\rfloor$). Using Lemma 2 they are Monge matrices as well. Consequently $\mathbf{A} = \mathbf{A}^M$. □

Equivalent conditions for a fuzzy interval matrix to be possibly robust were proved in [10]. However, the resulting matrix of the described algorithm with computational complexity $O(n^5)$ need not to have the Monge property ([7]). Due to Lemma 3 we need not to check the Monge property of the resulting matrix. However, we will prove equivalent conditions for possible robustness of an interval matrix of type Γ which guaranty an more effective algorithm with computational complexity $O(n^2)$ to check the possible robustness. Moreover, there is no polynomial algorithm for checking the universal robustness of interval matrices in fuzzy algebra. We will prove equivalent conditions for an interval matrix of type Γ to be universally robust and derive an algorithm with computational complexity $O(n^2)$ to check universal robustness.
Let us denote the set of all inputs by $H = \{\overline{a}_{ij};\ i,j \in N\} \cup \{\underline{a}_{ij};\ i,j \in N\}$.

Possible robustness

Theorem 4. *An interval matrix* $\mathbf{A} = [\underline{A}, \overline{A}]$ *of type* Γ *is possibly robust if and only if* $\sum_j^{\oplus} \underline{a}_{\gamma j} \otimes \underline{a}_{j\gamma} \leq \overline{a}_{\gamma\gamma}$.

Proof. First, we prove the sufficient condition. Let $\sum_j^{\oplus} \underline{a}_{\gamma j} \otimes \underline{a}_{j\gamma} \leq \overline{a}_{\gamma\gamma}$. We shall find a robust matrix $A^\star \in \mathbf{A}$. We set

$$a^\star_{ij} = \begin{cases} \overline{a}_{\gamma\gamma} & \text{if } i = j = \gamma, \\ \underline{a}_{ij} & \text{otherwise.} \end{cases} \tag{3}$$

We shall prove that $A^\star$ is robust. For each $h \in H$ the threshold digraph $G(A^\star, h)$ contains loop on node γ, hence by Remark 2 and Theorem 1 is $A^\star$ robust.
For the converse implication let us assume that $\mathbf{A}$ is possibly robust. Consequently there is a matrix $A \in \mathbf{A}$ which is robust. Let $\sum_j^{\oplus} \underline{a}_{\gamma j} \otimes \underline{a}_{j\gamma} > \overline{a}_{\gamma\gamma}$. Then for every matrix $A \in \mathbf{A}$ the threshold digraph $G(A, h)$ for $h = \sum_j^{\oplus} \underline{a}_{\gamma j} \otimes \underline{a}_{j\gamma}$ contains at least one cycle of length two but no loop on node γ. This implies using Remark 2 and Theorem 1 that A is not robust, what is in contradiction with the existence of a robust matrix in $\mathbf{A}$. Thus $\sum_j^{\oplus} \underline{a}_{\gamma j} \otimes \underline{a}_{j\gamma} \leq \overline{a}_{\gamma\gamma}$. □

Theorem 5. *There is an algorithm with computational complexity* $O(n^2)$ *for verifying the type of an interval matrix* $\mathbf{A} = [\underline{A}, \overline{A}]$ *and checking the possible robustness.*

Proof. Due to Theorem 3 verifying that a matrix is of type Γ takes $O(n^2)$ time. We shall check the lower bound matrix $\underline{A}$ and the upper bound matrix $\overline{A}$. To check the inequality $\sum_j^{\oplus} \underline{a}_{\gamma j} \otimes \underline{a}_{j\gamma} \leq \overline{a}_{\gamma\gamma}$ takes $O(n)$ time, hence the entire computational complexity is $O(n^2)$. □

In following two examples are the considered fuzzy matrices defined over the set $B = [0, 10]$. We illustrate the possibly non-robust and the possibly robust case for interval matrices of type $\lceil\Gamma$ using Theorem 4.

Example 1. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A}$, $\overline{A} \in B(5, 5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 4 & 1 & 1 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{pmatrix}, \ \overline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 4 & 2 & 3 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{pmatrix}. \qquad A^\star = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 4 & 1 & 1 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{pmatrix}$$

Both $\underline{A}$ and $\overline{A}$ are matrices of type $\lceil\Gamma$ with $\gamma = 2$. Since $\sum_j^{\oplus} \underline{a}_{\gamma j} \otimes \underline{a}_{j\gamma} = 3 \leq \overline{a}_{\gamma\gamma} = 3$ the considered interval matrix is possibly robust, hence there is a matrix $A^\star$ defined by (3) which is robust, and indeed the threshold digraph $G(A^\star, 4)$ is acyclic and $G(A^\star, h)$ for $h < 4$ contains loop on node $\gamma = 2$.

Example 2. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A}$, $\overline{A} \in B(5, 5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & 1 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{pmatrix}, \ \overline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 4 & 2 & 3 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{pmatrix}.$$

Both $\underline{A}$ and $\overline{A}$ are matrices of type $\lceil\Gamma$ with $\gamma = 2$ again. Since $\sum_j^{\oplus} \underline{a}_{\gamma j} \otimes \underline{a}_{j\gamma} = 3 > \overline{a}_{\gamma\gamma} = 2$ the considered interval matrix is possibly not robust, and indeed for every matrix $A \in \mathbf{A}$ the threshold digraph $G(A, 3)$ contains the cycle $c = (2, 3, 2)$ of length two but no loop on node $\gamma = 2$. Thus A is not robust.

Universal robustness

Theorem 6. *An interval matrix* $\mathbf{A} = [\underline{A}, \overline{A}]$ *of type* Γ *is universally robust if and only if the matrix* $\overline{A}$ *is robust and* $\underline{a}_{\gamma\gamma} \in \Big[\sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma},\ \overline{a}_{\gamma\gamma}\Big]$.

Proof. First, we prove the sufficient condition. Let $\overline{A}$ be robust. There are two possibilities. First, the trivial case, $G(\overline{A}, h)$ contains no cycle of length two for any $h > 0$. Consequently $G(A, h)$ contains no cycle of length two for any $h > 0$ and for every matrix $A \in \mathbf{A}$. Due to Remark 2 and Theorem 1 A is robust, hence $\mathbf{A}$ is universally robust. In this case $\sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma} = 0$, hence $\underline{a}_{\gamma\gamma} \in [0,\ \overline{a}_{\gamma\gamma}]$. Second, $G(\overline{A}, h)$ contains at least one cycle of length two for some $h > 0$, i.e. there is at least one $j \in N$ such that $\overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma} \geq h$. By Theorem 2 this implies $\overline{a}_{\gamma\gamma} \geq h$ and $G(\overline{A}, h)$ contains loop on node γ. Since $\underline{a}_{\gamma\gamma} \in \Big[\sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma},\ \overline{a}_{\gamma\gamma}\Big]$ the threshold digraph $G(A, h)$ for all matrices $A \in \mathbf{A}$ contains the loop on node γ for all $h \in H$. Thus by Theorem 1 and Remark 2 A is robust, hence $\mathbf{A}$ is universally robust.
Now, we prove the necessary condition. Let $\mathbf{A}$ be universally robust. Trivially $\overline{A}$ is robust. We shall prove that $\underline{a}_{\gamma\gamma} \in \Big[\sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma},\ \overline{a}_{\gamma\gamma}\Big]$. Let $\underline{a}_{\gamma\gamma} < \sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma}$. Then there is a matrix $A^\star \in \mathbf{A}$ defined as follows

$$a^\star_{ij} = \begin{cases} \underline{a}_{\gamma\gamma} & \text{if } i = j = \gamma, \\ \overline{a}_{ij} & \text{otherwise.} \end{cases} \tag{4}$$

We shall prove that $A^\star$ is not robust. Let us consider the threshold digraph $G(A^\star, h)$ for $h = \sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma}$. The digraph contains a non-trivial strongly connected component with a cycle of length two connecting node γ with the node k for which $\overline{a}_{\gamma k} \otimes \overline{a}_{k\gamma} = \sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma}$. Since $a^\star_{\gamma\gamma} = \underline{a}_{\gamma\gamma} < \sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma}$ the component is without a loop. Thus by Theorem 1 is the matrix $A^\star$ not robust. What is a contradiction with universal robustness of $\mathbf{A}$. □

Theorem 7. *There is an algorithm with computational complexity* $O(n^2)$ *for verifying the type of an interval matrix* $\mathbf{A} = [\underline{A}, \overline{A}]$ *and checking the universal robustness.*

Proof. Due to Theorem 3 verifying that a matrix is of type Γ takes $O(n^2)$ time. We shall check the lower bound matrix $\underline{A}$ and the upper bound matrix $\overline{A}$. To find the required interval and check whether $\underline{a}_{\gamma\gamma} \in \Big[\sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma},\ \overline{a}_{\gamma\gamma}\Big]$ takes $O(n)$ time, hence the entire computational complexity is $O(n^2)$. □

In following two examples are the considered fuzzy matrices defined over the set $B = [0, 10]$. We illustrate the universally robust and the universally non-robust case for interval matrices of type $\lceil\Gamma$ using Theorem 6.

Example 3. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A},\ \overline{A} \in B(5, 5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \end{pmatrix},\ \overline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 5 & 3 & 6 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \end{pmatrix}. \qquad A^\star = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 5 & 3 & 6 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \end{pmatrix}$$

Both $\underline{A}$ and $\overline{A}$ are matrices of type $\lceil\Gamma$ with $\gamma = 2$. Moreover $\overline{a}_{ij} \leq \overline{a}_{\gamma\gamma}$ for $i,\ j \in N$ implies that the matrix $\overline{A}$ is robust. Since $\underline{a}_{\gamma\gamma} = 3 < \sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma} = 5$ there is a matrix $A^\star$ defined by (4) which is not robust, namely the threshold digraph $G(A^\star, 5)$ contains one non-trivial strongly connected component, but without a loop (see Figure 1). Consequently the considered interval matrix $\mathbf{A}$ is not universally robust.

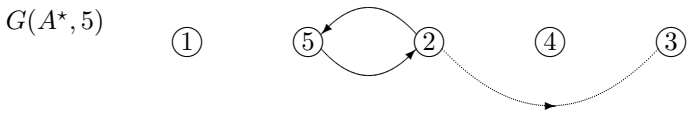


Figure 1 Universal robustness in non-robust case

Example 4. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A}$, $\overline{A} \in B(5,5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & k & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \end{pmatrix}, \ \overline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 5 & 3 & 6 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \end{pmatrix}.$$

Both $\underline{A}$ and $\overline{A}$ are matrices of type $\lceil\Gamma$ with $\gamma = 2$ again. Moreover $\overline{a}_{ij} \leq \overline{a}_{\gamma\gamma}$ for $i,\ j \in N$ implies that the matrix $\overline{A}$ is robust. Since $\sum_j^{\oplus} \overline{a}_{\gamma j} \otimes \overline{a}_{j\gamma} = 5$ and $\overline{a}_{\gamma\gamma} = 10$ for arbitrary $k \in [5, 10]$ contains for $A \in \mathbf{A}$, the threshold digraph $G(A, h)$ for each $h \in H$ exactly one non-trivial strongly connected component with a loop on node $\gamma = 2$. Thus A is robust and consequently $\mathbf{A}$ is universally robust.

5 Conclusion

The aim of this paper is to find equivalent conditions for robustness of Monge fuzzy matrices of type Γ with fixed data and equivalent conditions for possible robustness and universal robustness of interval Monge fuzzy matrices of type Γ.

The theory of matrix robustness can reflect the following economic application. We consider evaluation of projects in a company. The projects are characterized by different properties. The level of each property i is described by value x_i, influenced by all properties x_j. The influence is represented by a factor a_{ij}. The vector x at time $r+1$ can be computed by equation $x(r+1) = A \otimes x(r)$. The robustness of the matrix A is related to the solution of the eigenproblem $A \otimes x(r) = x(r)$, where the eigenvector $x(r)$ represents the steady state.

References

[1] Burkard, R. E., Klinz, B. and Rudolf, R.: Perspectives of Monge properties in optimization, *DAM*, Volume **70** (1996), 95–161.

[2] Fiedler, M., Nedoma, J., Ramík, J., Rohn, J. and Zimmermann, K.: Linear Optimization Problems with Inexact Data. Springer–Verlag, Berlin 2006.

[3] Gavalec, M., Ramík, J., and Zimmermann, K.: Decision Making and Optimization. Springer International Publishing Switzerland 2015.

[4] Gavalec, M. and Plavka, J.: An $O(n^2)$ algorithm for maximum cycle mean of Monge matrices in max algebra, *DAM*, Volume **127** (2003), 651–656.

[5] Gavalec, M. and Zimmermann, K.: Classification of solutions to systems of two-sided equations with interval coefficients, *Inter. J. of Pure and Applied Math.* Volume **45** (2008), 533–542.

[6] Molnárová, M.: Robustness of (0–1) Monge fuzzy matrices, In: *Proceedings of* 31^{st} *Int. Conference Mathematical Methods in Economics 2013*, Jihlava, 2013, 636–641.

[7] Molnárová, M.: Robustness of Monge matrices in fuzzy algebra, In: *Proceedings of* 32^{nd} *Int. Conference Mathematical Methods in Economics 2014*, Olomouc, 2014, 679–684.

[8] Molnárová, M.: Robustness of Monge fuzzy matrices with inexact data, In: *Proceedings of* 33^{rd} *Int. Conference Mathematical Methods in Economics 2015*, Cheb, 2015, 566–571.

[9] Molnárová, M.: Possible and universal robustness of Monge fuzzy matrices, In: *Proceedings of* 34^{th} *Int. Conference Mathematical Methods in Economics 2016*, Liberec, 2016, 582–587.

[10] Molnárová, M., Myšková, H. and Plavka, J.: The robustness of interval fuzzy matrices, *DAM*, Volume **438** (2013), 3350–3364.

[11] Plavka, J.: On the $O(n^3)$ algorithm for checking the strong robustness of interval fuzzy matrices, *DAM*, Volume **160** (2012), 640–647.

[12] Plavka, J. and Sergeev, S.: Characterizing matrices with X-simple image eigenspace in max-min algebra, *Kybernetika*, Volume **52** Issue 4 (2016), 497–513.

[13] Plavka, J. and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *DAM*, Volume **159** Issue 5 (2011), 381–388.

Can shifts in public debt structure influence economic growth of advanced economies?

Martin Murín[1]

Abstract. Since the Great Recession, dynamics of public debt have forced many politicians to choose whether they will boost the economic growth or they will deal with a high public debt to GDP ratio. Measures able to lower the debt and increase the economic growth simultaneously seems to be difficult to find. Following the most intuitional case of the consolidation when fiscal authority increases taxation or cuts public spending could not lead to better perform. If the debt declines slower than the GDP, the debt quota will rise anyway. From very simple example one can see an interaction of the public debt and the economic performance especially during economic downturns. Then, it is important to look for measures that can enhance the growth and are, at least, the debt neutral. Therefore, the aim of the contribution is to examine whether the level neutral change in structure of the public debt could have desirable effect on the economic growth of advanced economies. There are some theoretical linkages which suggest that the debt structure can influence the growth rates. Arnold's et al. [2] methodology enables to estimate effect of internal debt structure change which has no direct effect on the level of public debt. The dataset covers period from 2000 to 2015 of all OECD countries of which data are available for the econometric analysis (up to 24). Hence the panel data regression is employed to estimate results and 3 debt decompositions of the public debt are made. Namely, i) by the original maturity of debt instruments, ii) by the currency which is used to issue debt, iii) by the residency of creditors. The results suggest that the most important shift seems to be from domestic to foreign creditors which is associated with higher economic growth rates. Similar but not as strong effects is the shift from domestic to foreign currency and finally the shift from long term debt to short term debt instruments.

Keywords: economic growth, sovereign debt crisis, public debt, structure of public debt, panel data regression

JEL Classification: O47, H63, C23,
AMS Classification: 62-07

1 Introduction

In 2009 the sovereign debt crisis appeared in the EU. Since then the public debt has been held attention among economists. From the field of economic growth and public debt studies one must mention Reinhart and Rogoff [16] contribution which has led to much more intensive discussion even though the contribution has several mistakes, see Herndon et al. [10]. Reinhart and Rogoff show that from certain level of the public debt could be harmful to the economic growth. On the other hand, economic performance influences the level and the dynamic of the public debt. It is much more obvious if the ratio of debt to GDP is taking as an indicator of indebtedness. If one assumes the fiscal deficit to correspond to the debt difference than the public debt can be increased by means of automatic stabilizers in time of economic downturns and the debt quota tends to increase even more. The causality goes from the public debt to the economic growth is unambiguous, see Panizza and Presbitero [15]. The aim of this paper is to investigate whether the change in the public debt structure can have an impact on the economic growth of advanced economies. To do this, the Arnold et al. [2] methodology is used. However, they use this methodology to evaluated effects of tax shifts which are neutral to the level of the tax quota. Therefore, in this contribution, shifts in public debt with neutral effect on the level of public debt are empirically analyzed. The sample consist of 24 OECD countries in period from 2000 to 2015. Hence the panel data regression is employed to estimate result. Debt structure is divided by 3 criteria. Namely, i) by the original maturity of debt instruments, ii) by the residency of creditors, iii) by the currency which is used to issue debt. The results suggest that the most important shift seems to be from domestic to foreign creditors which is associated with the higher economic growth rates. Similar but not as strong effects is obtained from the shift from domestic to foreign currency and from the shift goes from long term debt instruments to short term debt instruments.

[1] Masaryk university, department of Economics, Faculty of Economics and Administration, Lipová 41/a, 60200 Brno, Czech Republic, martin.murin@mail.muni.cz.

2 Theoretical impact of public debt on economic growth

First of all, some theoretical background needs to be set to learn the way how the public debt fits into the economic growth model. It is followed Teles and Mussolini [19] which is an OLG model of the economic growth. This model extends Brauninger [6] by the approach to the government expenditure. Some share of the expenditure is productive. It means the productive expenditure has positive impact on the marginal product. In this setting, it is not look for the optimal size of the productive government expenditure as it is done by Bacchiocchi et al. [3] and there is no monetary sector as it is in Adam and Bevan [1]. There are three sectors in the economy.

Households

Households are consisted of two generations (0,1), old and young. Every is an inner homogeneous group. Utility function of agent born in period t = 0, 1 is given by eq. (1).

$$U = \ln c_t^t + \beta \ln c_{t+1}^t \tag{1}$$

In (1) c_j^t is a consumption in time j of an agent born in the period t. The parameter β shows agent preferences of the future consumption over the current consumption. It is assumed that β is greater than 0. This secure that there is some desire for the future consumption. Old generation doesn't work. Members are owners of the initial capital k_0. Young generation represents labor force and its budget constrain is as follows:

$$c_t^t + s_t^t \leq (1-\tau) w_t \tag{2}$$

$$c_{t+1}^t \leq (1 + r_{t+1}) s_t^t \tag{3}$$

where w_t is a wage, r_{t+1} is an interest rate, s_t^t is a saving and τ is an exogenous tax rate. One must note that in the period t+1 young generation become old, hence in this period there is no wage and the maximal amount of consumption in the period t+1 is given by the amount of saving from the previous period and the real interest rate. Household uses savings to buy private capital or government bonds.

$$s_t^t = k_{t+1} + d_{t+1} \tag{4}$$

Firms

The production of the economy is given by the representative firm, which, in correspondence with neoclassical theory, maximizes its profit. The production function is described by the eq. (5).

$$y_t = A z_t^{1-\alpha} k_t^{\alpha} l_t^{1-\alpha} \tag{5}$$

where y_t is an output, A is a level of technology, z_t is an amount of the productive government expenditure, l_t is a labor force which is, for simplicity, created by the young generation, l_t =1, k_t is a capital stock which follows the invest function (6).

$$k_{t+1} = (1-\delta) k_t + i_t \tag{6}$$

where the future capital stock depends on a rate of amortization δ, which is equal to 1 for simplicity. Therefore, the future capital stock is equal to an investment i_t. Government is represented by z_t in the production function. It is given as fallows.

$$z_t = x y_t \tag{7}$$

By substitution (7) into (5) it is gotten:

$$y_t = A^{1/\alpha} x^{(1-\alpha)/\alpha} k_t \tag{8}$$

From this relation stems that there is an externality given by the government productive expenditure which means that the output of representative firm can be higher without a change in the private capital stock.

Government

Government is facing its budget constrain. In the eq. (9) government has productive expenditure x and unproductive expenditure g. Both are dependent on the level of output. Government revenue is given by the tax rate τ and level of wage w_t. There is no other source of the government revenue.

$$d_{t+1} - d_t = r_t d_t + \left[(g + x)y_t - \tau w_t\right] \tag{9}$$

Change of the public debt in the future period is given by a difference between government revenue and government expenditure in the current period. Relation in parentheses is a primary deficit. The debt service costs are represented by the amount of the debt in the current period multiplied by the interest rate. An initial level of the debt d_0 is equal to 0, like in Adam and Bevan [1].

Market equilibrium

An equilibrium is given by maximizing the utility function of young generation (1) and the product function (5). It is needed to put constrain on the economy.

$$y_t = c_t^t + c_t^{t-1} + k_{t+1} + (g + x)y_t \tag{10}$$

The following relation (11) is obtained by maximizing utility for given w_t and r_{t+1}.

$$s_t^t = \beta(1-\tau)w_t / (1+\beta) \tag{11}$$

The saving is positively dependent on preference toward the future consumption and the level of wage after tax. Optimal behavior of representative firm is given by following relation:

$$\partial y_t / \partial k_t = \alpha A z_t^{1-\alpha} k_t^{\alpha-1} = \alpha A^{1/\alpha} x^{(1-\alpha)/\alpha} = r_t \tag{12}$$

The marginal product of the private capital is equal to the real interest rate. There is a perfect competition which means there is no economic profit, see the eq. (13).

$$y_t - r_t k_t = w_t \tag{13}$$

Hence wage of young generation is given by the eq. (14).

$$w_t = (1-\alpha)A^{1/\alpha} x^{(1-\alpha)/\alpha} k_t \tag{14}$$

Now, it can be said, that if government increases the productive expenditure it increases the marginal product of the private capital and the real interest rate, which leads to the higher debt cost (9). On the other hand, it increases the wage of young generation and using (11) it seems to increase the saving.

Growth of capital stock

As suggested by Teles and Mussolini [19] the economic growth rate can be expressed by using a growth of private capital stock. Equation (15) shows case, when government decides about the tax rate and the deficit is endogenous. It is possible to fix deficit and then consider the tax rate as endogenous, however, this situation in public finance could lead to vary taxation often.

$$\frac{k_{t+1} - k_t}{k_t} = \frac{\beta}{1+\beta}(1-\tau)(1-\alpha)A^{1/\alpha} x^{(1-\alpha)/\alpha} - \left[g + x - (1-\alpha)\tau\right]A^{1/\alpha} x^{(1-\alpha)/\alpha} - \left(1 + \alpha A^{1/\alpha} x^{(1-\alpha)/\alpha}\right)\frac{d_t}{k_t} - 1 \tag{15}$$

From (15) stems that the growth rate of capital accumulation is positively dependent on the level of agents' propensity to the future consumption and the net marginal product of the private capital which is in correspondence with Barro [5]. Second part of the eq. (15) is a negative effect of the level of the primary deficit similarly to Brauninger [6]. Last part of the eq. (15) is an effect resulting from the ratio of public debt cost to capital accumulation. It means, if the public debt service cost is high then the economic growth is slow. This negative impact is higher if the debt stock is higher, or the interest rate is higher or the private capital is low. The result of the model is according to Teles and Mussolini [19] similar to the PAYG system.

3 Empirical analysis

From theoretical model, the importance of the public debt cost can be understood. Higher the cost lower the economic growth. From this point of view the structure of the public debt could play a significant role to influence the rate of real GDP growth in the long term.

3.1 Methodology

This contribution focus on the question how does the shift in the public debt structure influences the economic growth? It is common to use panel data regression to estimate an endogenous growth model, for example see Vráblíková [20] or Zimčík [21]. In compliance with the aim of this paper the most intuitive approach is to divide public debt into categories mentioned within the introduction. In previous research, see Murín [14], it is used the omitted variable approach developed by Kneller et al. [11]. Authors Arnold et al. [2] modify the omitted variable approach to review an effect of the tax shift on the economic performance with no direct impact on the tax quota. Murín [13] uses this approach to investigate effects of the level neutral shifts in public debt structure on economic growth of EU 14 countries. This approach allows to perform relatively easy econometric exercise. Equation (16) shows the econometric model.

$$g = X\beta + DEBT\gamma + \varepsilon \tag{16}$$

where g is dependent variable, X is a set of control variables and $DEBT$ is a set of variables of interest, in this contribution a set of all debt structure variables. The approach can be described as follows. One put the level of public debt quota into the set of control variables to adjust estimates of its impact on the economic growth. The set of variables of research interest are calculated as a share of the whole public debt. If one share is omitted, all estimated coefficients on shares leaved in the regression can be interpreted as a resulted effect of the structure shift towards this share with the opposite change of the debt share which is omitted. This is very similar approach to Kneller et al. [11] omitted variable approach, however, the crucial difference is that the level of total variable is given into the set of control variable and variables of interest are calculated as shares of total variable. Arnold et al. [2] study the tax shift this way.

Data description

As it is mentioned above, in similar contributions the endogenous model is estimated. In these models, variables are divided into two categories. The first is the set of control variables and the second is set of variables of research interest. Table 1 presents all variables, their labels and verbal description.

Label	Verbal description
Growth	Year-to-year growth rate of real GDP per capita, constant prices, constant exchange rates
Invest	Gross fixed capital formation, percentage of GDP, current prices, in logarithm
Open	Percentage of export + import to GDP, current prices, in logarithm
Tech	Thousand patents per inhabitant, in logarithm
Gov	Percentage of total public expenditure to GDP, general government, current prices, in logarithm
Debt	Gross public debt-to-GDP ratio, current prices (in all next: c. p.)
i)	
Short	Short-term debt by original maturity, share of gross public debt, c. p.
Long	Long-term debt by original maturity, share of gross public debt, c. p.
ii)	
Curdom	Debt in domestic currency, share of gross public debt, c. p.
Curfor	Debt in foreign currency, share of gross public debt, c. p.
iii)	
Crefor	Debt held by external creditors, share of gross public debt, c. p.
Credom	Debt held by domestic creditors, share of gross public debt, c. p.

Table 1 Variables description

Table 1 shows all variables and their verbal description used in this contribution. The dependent variable is *Growth*, the y-o-y growth rate of the real GDP per capita. The set of control variables consists of *Invest*, which approximates physical capital accumulation in line with Solow [18], *Open*, which is used to control estimates for an effect of openness of the economy. If there is no such variable, all coefficients on debt categories tent to be more significant, see Murín, [13]. Variable *Tech* is used to approximate technological progress or at least the innovation intensity in line with Romer [17]. If one wants to study the effect of the public debt structure on the

economic growth it is quite reasonable to control estimates of the size of government. *Gov* is employed to do it so. Finally, the level of public debt, *Debt*, is incorporated into the set of control variable.

In Table 1 the set of variables of the research interest consists of labels i) the origin maturity of debt instruments, ii) the currency which is used to issue debt, iii) the residency of creditors. The sample consist of OECD countries with satisfying data availability on public debt structure. Namely they are Australia, Austria, Canada, Estonia, Finland, France, Hungary, Ireland, Italy, Luxemburg, Portugal, Slovakia, Spain, Sweden, Great Britain and USA. Followed period is from 1998 to 2015. Every variable was drawn from open OECD.Stat database. The OECD captures debt structure on a quarterly periodicity. In the analytical part of this contribution, it is worked with data in yearly periodicity, as e. g Machová and Kotlán [12], etc. Hence, it was important to find a way how data on the debt structure should be used. It has been decided to take values of the last quarters of the year to get debt structure of that year, because the debt is a state value, this method seems to be valid. Therefore, for instance the value of last quarter of year 2015 represents the state of debt structure in year 2015.

Equation (17) and (18) presents eq. (16) which is got by using (16) and verbal description. Eq. (17) represents estimated model in the next section. Index *i* stands for the country and index *t* denotes the period. Index *k* indicates specific debt ratio D_r_{kit-1} (variables of the interest in table 1) and index *j* denotes the omitted debt ration from the estimate. Eq. (18) shows that error term contains a country specific effect. By using Hausman test estimates were performed with fixed effects, see Batlagi [4]. For detailed discussion on method of estimation, see e.g. Murín [14].

$$Growth_{it} = \beta_0 + \beta_1 Invest_{it} + \beta_2 Open_{it} + \beta_3 Tech_{it} + \beta_4 Gov_{it} + \beta_5 Debt_{it-1} + \sum_{k=1}^{K-1}(\gamma_k - \gamma_j) D\_r_{kit-1} + \varepsilon_{it} \tag{17}$$

$$\varepsilon_{it} = \nu_i + u_{it} \tag{18}$$

3.2 Results

Table 2 shows results of all estimates relevant to the scope of this contribution. Signs ***, ** and * following the estimated parameter stand for significance level of 1, 5 and 10 percent respectively. The number in the brackets are t-statistics. It needs to be emphasized that the set of control variables in table 2 is the most appropriate one from other specifications. Many other variables were tested which could approximate control variable. Especially in case of *Tech* and *Open*, there could be other options used. All control variables have anticipated effect and they are statistically significant. It is not the purpose of this contribution to concern about these variables. The aim of the contribution is to examine whether the level neutral change in the structure of public debt could have desirable effect on the economic growth of advanced economies.

From the point of view of the purpose the results suggest some interesting findings. Model 1 and 2 are dealing with shifts in the public debt from short term debt instruments to long term and vice-versa. Estimated parameters are significant at level 10 %, which is the lowest signification among others. From results stem that if debt structure shift goes from long-term to short-term instruments it is associated with higher economic growth. This could be explained by the model proposed above. If a yield curve is of normal shape it means that the long-term obligations should have higher interest rate than the short-term.

Models 3 and 4 are dealing with shifts in debt structure following the currency decomposition. There are a domestic and a foreign currency. Dell´Erba et al. [8] suggest that for the country it is much less risky to emit debt in the domestic currency. If the government cannot sell bond in the domestic currency even on the domestic market it is called the Hausman´s hypothesis of the origin sin. From results of model 3 and 4 it seems the hypothesis to be valid even for the economic growth. The shift from foreign currency to domestic currency is correlated with increase of the economic growth. The opposite shift leads to decrease of the economic growth.

Models 5 and 6 are devoted to the last debt decomposition. It deals with the residency of creditors. Since the 1950s, it has been discussion whether are better to use domestic or foreign sources to finance public debt, see Buchanan [7]. From results in table 2 stem that it is better to supply debt obligation to foreign creditors that to domestic ones. This result is in correspondence with Adam and Bevan [1] or Diamond [9].

Model	**1**	**2**	**3**	**4**	**5**	**6**
β_0	1,439***	1,440***	1,291***	1,290***	1,109***	1,112***
	(22,700)	(22,661)	(14,161)	(14,150)	(13,381)	(13,308)

Invest	15,371***	15,373***	11,421***	11,410***	10,700***	10,680***
	(5,316)	(5,331)	(3,315)	(3,310)	(3,907)	(3,878)
Open	11,866***	11,876***	11.,447***	11,446***	13,193***	13,166***
	(7,396)	(7,413)	(4,335)	(4,335)	(6,064)	(6,022)
Tech	3,010***	3,006***	1,981*	1,995*	3,482***	3,478***
	(3,626)	(3,615)	(1,630)	(1,643)	(2,753)	(2,742)
Gov	-18,35***	-18,37***	-15,15***	-15,16***	-17,73***	-17,70***
	(-5,500)	(-5.500)	(-3,409)	(-3,408)	(-4,308)	(-4,304)
Debt	-0,092***	-0,092***	-0,119***	-0,119***	-0,110***	-0,110***
	(-4,379)	(-4,381)	(-3,351)	(-3,352)	(-4,653)	(-4,639)
Short	0,053*					
	(1,824)					
Long		-0,055*				
		(-1,873)				
Curdom			0,054***			
			(3,211)			
Curfor				-0.054***		
				(-3,218)		
Credom					-0,092***	
					(-2,769)	
Crefor						0,090***
						(2,688)
Adj. R^2	0,681	0,681	0,626	0,626	0,671	0,669
No.	317	317	185	185	208	208

Table 2 Results of estimates of the effect of public debt structural shifts on economic growth

4 Conclusions

The aim of the contribution was to examine whether the level neutral change in the structure of public debt could have desirable effect on the economic growth of advanced economies. From the empirical point of view the panel data regression of 24 OECD countries in the period from 2000 to 2015 was used to estimates economic growth models with public debt and its structure variables. The decomposition of structure of public debt was performed: i) by the original maturity of debt instruments, ii) by the currency which are used to issue debt, iii) by the residency of creditors. Arnold et al. [2] approach was followed to observe effects of shifts in the public debt structure which don´t change level of the debt quota.

The results suggest that shifts in public debt structure are relevant to the economic growth of the researched sample. According to sizes of estimated parameters the most important shift is by the residency of creditors, followed by the currency and finally by the original maturity of debt instruments. All results can be related to the model presented in section 2. All shifts in the debt structure which are positively correlated with the rate of growth of the economy can be explained by means of their impact on the interest rate and then on the debt service cost. If there is a shift with additional pressure on debt cost it should lower the economic growth. Hence it is better for the growth to shift debt structure towards foreign creditors, sell obligations in domestic currency if possible, and use short-term obligations which requires to roll-over public debt. For better understanding of the theoretical model it would be interested to test the hypothesis of debt structure shift effect on public debt cost directly in future work.

Acknowledgements

This work was supported by funding of specific research at Faculty of Economics and Administration – project MUNI/A/1076/2016. This support is gratefully acknowledged.

References

[1] Adam, C. S. and Bevan, D. L.: Fiscal deficits and Growth in Developing Countries. *Journal of Public Economics* 89 (2005), 571-597.

[2] Arnold, J. M., Brys, B., Heady, Ch., Johansson, A., Schwellnus, C., and Vartia, L.: Tax Policy for Economic Recovery and Growth. The Economic Journal 121 (2011), 59-80.

[3] Bacchiocchi, E., Borghi, E. and Missale, A.: Public Investment under Fiscal Constraints. Fiscal Studies 32 (2011), 11-42.
[4] Baltagi, B. H.: Econometric Analysis of Panel Data, John Wiley & Sons Inc., New York, 2005.
[5] Barro, R. J.: Government Spending in A Simple Model of Endogenous Growth. The Journal of Political Economy 98 (1990), 103–125.
[6] Brauninger, M.: The Budget Deficit, Public Debt, and Endogenous Growth. Journal of Public Economic Theory 7 (2005), 827-840.
[7] Buchanan, J. M.: Public Principles of Public Debt: A Defense and Restatement. Richard D. Irwin, Homewood, 1958
[8] Dell'Erba, S., Hausmann, R. and Panizza, U: Debt Levels, Debt Composition, and Sovereign Spreads in Emerging and Advanced Economies. HKS Faculty. Working Paper RWP13-028, 2013.
[9] Diamond, P.: National Debt in a Neoclassical Growth Model. American Economic Review 55 (1965), 1126-1150.
[10] Herndon, T., Ash, M., Pollin, R.: Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff. Cambridge Journal of Economics 38 (2014), 257-279.
[11] Kneller, R., Bleaney, M. and Gemmell, N.: Fiscal Policy and Growth: Evidence from OECD Countries. Journal of Public Economics 74 (1999), 171–190.
[12] Machová, Z. and Kotlán, I.: Právní jistota – možný problém daňové politiky vyspělých zemí? Politická ekonomie 63 (2015), 833-846.
[13] Murín, M.: Public Debt and Economic Growth: An Empirical Analysis of the Structure Changes of Public Debt in the EU Countries. In: Proceeding from the 14th International Scientific Conference "Economic Policy in the European Union Member Countries" (Majerová, I. and Kotlánová, E., eds.), Silesian University in Opava, Karviná, 2016, 481-489.
[14] Murín, M.: Vplyv spôsobu tvorby fiškálneho deficitu na ekonomický rast. Politická ekonomie 64 (2016), 176-192.
[15] Panizza, U., and Presbitero, A. F.: Public Debt and Economic Growth: Is There a Causal Effect? Journal of Macroeconomics 41 (2014), 21–41.
[16] Reinhart, C. M. and Rogoff, K. S.: Growth in a Time of Debt", working paper [15639], National Bureau of Economic Research, Cambridge, MA, 2010.
[17] Romer, P. M.: Endogenous Technological Change. Journal of Political Economy 98 (1990), S71-S102.
[18] Solow, R. M.: A Contribution to the Theory of Economic Growth. The Quarterly Journal of Economics 70 (1956), 65-94.
[19] Teles, V. K. and Mussolini, C. C.: Public Debt and the Limits of Fiscal Policy to Increase Economic Growth. European Economic Review 66 (2014), 1-15.
[20] Vráblíková, V.: Vliv nepřímích daní na dlouhodobý ekonomický růst. Politická ekonomie 64 (2016), 145-160.
[21] Zimčík, P.: Velikost veřejného sektoru a ekonomický růst. Politická ekonomie 64 (2016), 439-450.

Interval Max-min Matrix Equations with Bounded Solution

Helena Myšková[1]

Abstract. Max-min algebra is an algebraic structure, in which classical addition and multiplication are replaced by maximum and minimum, respectively. Max-min (fuzzy) equations have found a broad area of applications in causal models which emphasize relationships between input and output variables. They are used in diagnosis models or models of non-deterministic systems. Many practical situations can be described using max-min matrix equations. It often happens that a max-min matrix equation with exact data is unsolvable. Therefore, we replace matrix elements with intervals of possible values. In this way, we obtain an interval matrix equation. We shall deal with interval max-min matrix equations with bounded solution, in which a solution has to be from the given interval of possible values. We can define several types of solvability of interval max-min matrix equations with bounded solution. In this paper, we shall deal with four of them. We prove the necessary and sufficient conditions which can be verified in polynomial time.

Keywords: max-min algebra, interval matrix equation, strong solvability, universal solvability

JEL classification: C02
AMS classification: 15A18; 15A80; 65G30

1 Motivation

Max-min equations have found a broad area of applications in causal models which emphasize relationships between input and output variables. They are used for instance in diagnosis models [6], [8], [9]. The solution of the max-min relational equation of the form $A \otimes x = b$, where A is a matrix, b and x are vectors of suitable dimensions and classical addition and multiplication operations are replaced by maximum and minimum, provides a maximal set of symptoms that produce the given fault.

The solvability of the systems of max-min linear equations is well reviewed. In this paper, we shall deal with the solvability of max-min matrix equations of the form $A \otimes X \otimes C = B$, where A, B, and C are given matrices of suitable sizes and X is an unknown matrix. In the following example we will show one of possible applications.

Example 1. Let us consider a situation, in which passengers from places P_1, P_2, P_3, and P_4 want to transfer to holiday destinations D_1, D_2, and D_3. Different transportation means provide transporting passengers from places P_1, P_2, P_3, and P_4 to airport terminals T_1 and T_2 (See Figure 1). We assume that the connection between P_i and T_l is possible only via one of the check points Q_1, Q_2, and Q_3. On Figure 1 there is an arrow $(P_i\, Q_j)$ if there exists the road from P_i to Q_j and there is an arrow $(T_l\, D_k)$ if terminal T_l handles the passengers traveling to destination D_k ($i = 1, 2, 3, 4$, $j = 1, 2, 3$, $k = 1, 2, 3$, $l = 1, 2$). The symbols along arrows represent the capacities of the corresponding connections. Denote by a_{ij} (c_{lk}) the capacity of the road from P_i to Q_j (from T_l to D_k). If Q_j is linked with T_l by a road with the capacity x_{jl}, then the capacity of the connection between P_i and D_k via Q_j using terminal T_l is equal to $\min\{a_{ij}, x_{jl}, c_{lk}\}$.

Suppose that the number of passengers traveling from place P_i to destination D_k is denoted by b_{ik}. To ensure the transportation for all passengers from P_1 to their destinations the following equations have to be satisfied:

$$\max\Big\{\min\{a_{11}, x_{11}, c_{11}\}, \min\{a_{12}, x_{21}, c_{11}\}\Big\} = b_{11},$$
$$\max\Big\{\min\{a_{11}, x_{11}, c_{12}\}, \min\{a_{12}, x_{21}, c_{12}\}, \min\{a_{12}, x_{22}, c_{22}\}\Big\} = b_{12},$$
$$\max\Big\{\min\{a_{11}, x_{12}, c_{23}\}, \min\{a_{11}, x_{11}, c_{13}\}, \min\{a_{12}, x_{21}, c_{13}\}, \min\{a_{12}, x_{22}, c_{23}\}\Big\} = b_{13}.$$

The similar equalities must be satisfied to ensure the transportation for all passengers from P_2, P_3 and P_4 to their destinations.

In general, suppose that there are m places $P_1, P_2, \ldots, P_m$, n transfer points $Q_1, Q_2, \ldots, Q_n$, s terminals $T_1, T_2, \ldots, T_s$, and r destinations $D_1, D_2, \ldots, D_r$. If there is no road from P_i to Q_j (from T_l to D_k), we put

[1]Technical University in Košice/ Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Němcovej 32, 042 00 Košice, Slovak Republic, helena.myskova@tuke.sk

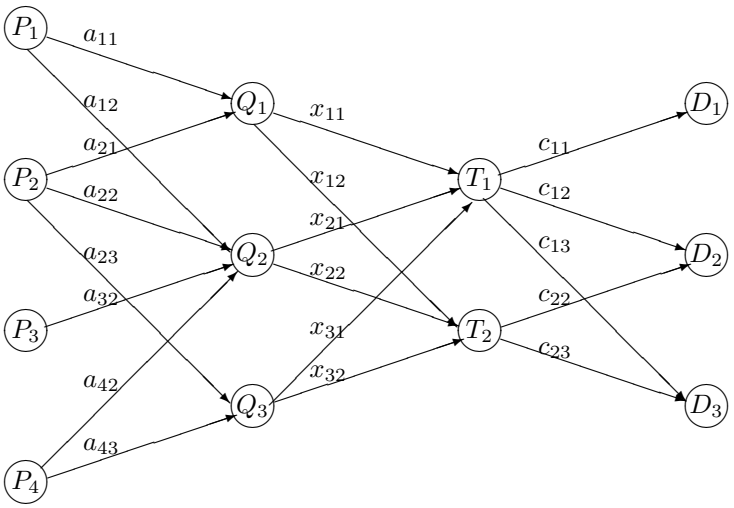


Figure 1 Transportation system

$a_{ij} = 0$ ($c_{lk} = 0$). Our task is to choose the appropriate capacities x_{jl} for any $j = 1, 2, \ldots, n$, and for any $l = 1, 2, \ldots, s$ such that the maximum capacity of the road from P_i to D_k is equal to a given number b_{ik} for any $i = 1, 2, \ldots, m$ and for any $k = 1, 2, \ldots, r$, i. e.,

$$\max_{j \in N,\, l \in S} \min\{a_{ij}, x_{jl}, c_{lk}\} = b_{ik}. \tag{1}$$

2 Preliminaries

Max-min algebra is the triple $(\mathcal{I}, \oplus, \otimes)$, where $\mathcal{I} = [O, I]$ is a linear ordered set with the least element O, the greatest element I, and two binary operations $a \oplus b = \max\{a, b\}$ and $a \otimes b = \min\{a, b\}$.

Denote by M, N, R, and S the index sets $\{1, 2, \ldots, m\}$, $\{1, 2, \ldots, n\}$, $\{1, 2, \ldots, r\}$, and $\{1, 2, \ldots, s\}$, respectively. The set of all $m \times n$ matrices over $\mathcal{I}$ is denoted by $\mathcal{I}(m, n)$ and the set of all column n vectors over $\mathcal{I}$ we denote by $\mathcal{I}(n)$.
Operations $\oplus$ and $\otimes$ are extended to matrices and vectors in the same way as in the classical algebra. For any $A, C \in \mathcal{I}(m, n)$, we write $A \leq C$ if $a_{ij} \leq c_{ij}$ for all $i \in M,\ j \in N$. The *monotonicity of* $\otimes$ means that for each $A, C \in \mathcal{I}(m, n),\ A \leq C$ and for each $B,\ D \in \mathcal{I}(n, s),\ B \leq D$ the inequality $A \otimes B \leq C \otimes D$ holds true.

3 Max-min Matrix Equations

Let $A \in \mathcal{I}(m, n)$, $B \in \mathcal{I}(m, r)$, and $C \in \mathcal{I}(s, r)$ be matrices with elements a_{ij}, b_{ik}, and c_{lk}, respectively. We shall write system of equalities (1) from Example 1 in the matrix form

$$A \otimes X \otimes C = B. \tag{2}$$

Denote by $X^*(A, B, C) = \big(x^*_{jl}(A, B, C)\big)$ the matrix defined as follows

$$x^*_{jl}(A, B, C) = \min_{k \in R}\{x^*_j(A \otimes c_{lk}, B_k)\}. \tag{3}$$

We call the matrix $X^*(A, B, C)$ a *principal matrix solution* of (2). The following theorem expresses the properties of $X^*(A, B, C)$ and gives the necessary and sufficient condition for the solvability of (2).

Theorem 1 (Draženská–Myšková [1]). *Let $A \in \mathcal{I}(m, n)$, $B \in \mathcal{I}(m, r)$ and $C \in \mathcal{I}(m, n)$.*
(i) *If $A \otimes X \otimes C = B$ for $X \in \mathcal{I}(n, s)$, then $X \leq X^*(A, B, C)$.*
(ii) $A \otimes X^*(A, B, C) \otimes C \leq B$.
(iii)*The matrix equation $A \otimes X \otimes C = B$ is solvable if and only if $X^*(A, B, C)$ is its solution.*

In practice, it happens quite often in modeling real situations that the obtained system turns out to be unsolvable. One of possible methods of restoring the solvability is to replace the exact input values by intervals of possible values. The result of the substitution is so-called interval max-min matrix equation. In this paper, we shall deal with the solvability of interval max-min matrix equations with bounded solution.

4 Interval Matrix Equations

Similarly to [2, 4, 7], we define *interval matrices* $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ as follows:

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \mathcal{I}(m,n);\ \underline{A} \leq A \leq \overline{A} \right\}, \quad \boldsymbol{B} = [\underline{B}, \overline{B}] = \left\{ B \in \mathcal{I}(m,r);\ \underline{B} \leq B \leq \overline{B} \right\},$$

$$\boldsymbol{C} = [\underline{C}, \overline{C}] = \left\{ C \in \mathcal{I}(s,r);\ \underline{C} \leq C \leq \overline{C} \right\}, \quad \boldsymbol{X} = [\underline{X}, \overline{X}] = \left\{ X \in \mathcal{I}(n,s);\ \underline{X} \leq X \leq \overline{X} \right\}.$$

We shall use the notation

$$\boldsymbol{A} \otimes \boldsymbol{X} \otimes \boldsymbol{C} = \boldsymbol{B} \tag{4}$$

for the set of all matrix equations of the form (2) such that $A \in \boldsymbol{A},\ B \in \boldsymbol{B}$, and $C \in \boldsymbol{C}$ and a solution is required to be from $\boldsymbol{X}$. We call (5) an *interval max-min matrix equation with bounded solution*.

A special case of (4) is

$$\boldsymbol{A} \otimes X \otimes \boldsymbol{C} = \boldsymbol{B}, \tag{5}$$

where $X \in \mathcal{I}(n,s)$ can be arbitrary.

In [5], there were studied several solvability concepts for (5). In this paper, we shall deal with the solvability of (4). We define some types of solvability.

Definition 1. Interval max-min matrix equation of the form (4) is

- *strongly solvable* if for each $X \in \mathcal{I}(n,s)$, for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$ the product $A \otimes X \otimes C$ lies in $\boldsymbol{B}$;
- *X-universally solvable* if for each $X \in \mathcal{I}(n,s)$ there exists $B \in \boldsymbol{B}$ such that the equality $A \otimes X \otimes C = B$ holds for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$;
- *A-universally solvable* if for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$ there exists $B \in \boldsymbol{B}$ such that the equality $A \otimes X \otimes C = B$ holds for each $X \in \mathcal{I}(n,s)$;
- *universally solvable* if there exists $B \in \boldsymbol{B}$ such that the equality $A \otimes X \otimes C = B$ holds for each $A \in \boldsymbol{A}$, for each $C \in \boldsymbol{C}$ and for each $X \in \mathcal{I}(n,s)$.

Theorem 2. *Interval max-min matrix equation of the form* (4) *is strongly solvable if and only if the system of inequalities*

$$\underline{A} \otimes \underline{X} \otimes \underline{C} \geq \underline{B}, \tag{6}$$

$$\overline{A} \otimes \overline{X} \otimes \overline{C} \leq \overline{B} \tag{7}$$

is satisfied.

Proof. The strong solvability means that the system of inequalities $\underline{B} \leq A \otimes X \otimes C \leq \overline{B}$ holds for each $A \in \boldsymbol{A}$, $C \in \boldsymbol{C}$ and for each $X \in \boldsymbol{X}$. The left inequality holds for each $A \in \boldsymbol{A}$, $C \in \boldsymbol{C}$ and for each $X \in \boldsymbol{X}$ if and only if it holds for $A = \underline{A}$, $C = \underline{C}$ and $X = \underline{X}$, i.e., (6) is satisfied. Similarly, the validity of the right inequality is equivalent to (7). □

It follows from the definitions of the strong, X-universal and A-universal solvability that the strong solvability is a necessary condition for the X-universal solvability and the A-universal solvability, too. For the given indices $i \in M,\ j \in N$ denote the matrix $A^{(ij)} = (a_{uv}^{(ij)})$ where

$$(a_{uv}^{(ij)}) = \begin{cases} \overline{a}_{ij} & \text{for } i = u,\ j = v, \\ \underline{a}_{ij} & \text{otherwise.} \end{cases}$$

Lemma 3 (Myšková [3])**.** Let $A \in \boldsymbol{A}$. Then

$$A = \bigoplus_{i \in M, j \in N} a_{ij} \otimes A^{(ij)}. \tag{8}$$

It is easy to see that Lemma 3 can be used in the similar manner for any interval matrix.

Lemma 4. Interval max-min matrix equation (4) is X-universally solvable if and only if interval max-min matrix equation (4) is strongly solvable and the equality

$$\underline{A} \otimes X \otimes \underline{C} = \overline{A} \otimes X \otimes \overline{C} \tag{9}$$

is satisfied for each $X \in \boldsymbol{X}$.

Proof. The strong solvability implies that $\underline{A} \otimes X \otimes \underline{C} \geq \underline{A} \otimes \underline{X} \otimes \underline{C} \geq \underline{B}$ and $\overline{A} \otimes X \otimes \overline{C} \leq \overline{A} \otimes \overline{X} \otimes \overline{C} \leq \overline{B}$ which means that $\underline{A} \otimes X \otimes \underline{C} \in \boldsymbol{B}$ and $\overline{A} \otimes X \otimes \overline{C} \in \boldsymbol{B}$ for each $X \in \boldsymbol{X}$. Let $X \in \boldsymbol{X}$ be fixed. The existence of $B \in \boldsymbol{B}$ such that for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$ the equality $A \otimes X \otimes C = B$ is satisfied means that the products $A \otimes X \otimes C$ are equal to each other for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$, which is equivalent to (9). □

Lemma 4 does not provide an effective algorithm for testing the X-universal solvability.

Theorem 5. *Interval max-min matrix equation* (4) *is X-universally solvable if and only if interval max-min matrix equation* (4) *is strongly solvable and*

$$\underline{A} \otimes X^{(jl)} \otimes \underline{C} = \overline{A} \otimes X^{(jl)} \otimes \overline{C} \tag{10}$$

for each $j \in N$, $l \in S$.

Proof. Suppose that $\underline{A} \otimes X^{(jl)} \otimes \underline{C} = \overline{A} \otimes X^{(jl)} \otimes \overline{C}$ for each $j \in N$, $l \in S$ and interval max-min matrix equation (4) is strongly solvable. According to Lemma 3 we obtain

$$\underline{A} \otimes X \otimes \underline{C} = \underline{A} \otimes \Big(\bigoplus_{j \in N, l \in S} x_{jl} \otimes X^{(jl)} \Big) \otimes \underline{C} = \bigoplus_{j \in N, l \in S} x_{jl} \otimes (\underline{A} \otimes X^{(jl)} \otimes \underline{C}) =$$

$$\bigoplus_{j \in N, l \in S} x_{jl} \otimes (\overline{A} \otimes X^{(jl)} \otimes \overline{C}) = \overline{A} \otimes \Big(\bigoplus_{j \in N, l \in S} x_{jl} \otimes X^{(jl)} \Big) \otimes \overline{C} = \overline{A} \otimes X \otimes \overline{C}$$

for each $X \in \boldsymbol{X}$. By Lemma 4 interval system (4) is X-universally solvable.
The converse implication is trivial. □

Lemma 6. Interval max-min matrix equation (4) is A-universally solvable if and only if interval max-min matrix equation (4) is strongly solvable and

$$A \otimes \underline{X} \otimes C = A \otimes \overline{X} \otimes C \tag{11}$$

for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$.

Proof. The strong solvability ensures that and $A \otimes X \otimes C \in \boldsymbol{B}$ for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$. For fixed $A \in \boldsymbol{A}$ and for fixed $C \in \boldsymbol{C}$, the A-universal solvability means that the products $A \otimes X \otimes C$ are the same for each $X \in \boldsymbol{X}$, which is equivalent to (11). □

The following theorem provides the polynomially testable condition for the A-universal solvability.

Theorem 7. *Interval max-min matrix equation* (4) *is A-universally solvable if and only if interval max-min matrix equation* (4) *is strongly solvable and the equality*

$$A^{(ij)} \otimes \underline{X} \otimes C^{(lk)} = A^{(ij)} \otimes \overline{X} \otimes C^{(lk)} \tag{12}$$

holds for each $i \in M$, $j \in N$, $k \in R$, $l \in S$.

Proof. Suppose that equality (12) is satisfied for each $i \in M$, $j \in N$ and interval max-min matrix equation (4) is strongly solvable. According to Lemma 3 we have

$$A \otimes \underline{X} \otimes C = \Big(\bigoplus_{i,j} a_{ij} \otimes A^{(ij)} \Big) \otimes \underline{X} \otimes \Big(\bigoplus_{l,k} c_{lk} \otimes C^{(lk)} \Big) = \Big(\bigoplus_{i,j} a_{ij} \otimes A^{(ij)} \otimes \underline{X} \Big) \otimes \Big(\bigoplus_{l,k} c_{lk} \otimes C^{(lk)} \Big) =$$

$$\bigoplus_{i,j,k,l} \Big((a_{ij} \otimes c_{lk}) \otimes \big(A^{(ij)} \otimes \underline{X} \otimes C^{(lk)} \big) \Big) = \bigoplus_{i,j,k,l} \Big((a_{ij} \otimes c_{lk}) \otimes \big(A^{(ij)} \otimes \overline{X} \otimes C^{(lk)} \big) \Big) =$$

$$\Big(\bigoplus_{i,j} a_{ij} \otimes A^{(ij)} \Big) \otimes \overline{X} \otimes \Big(\bigoplus_{l,k} c_{lk} \otimes C^{(lk)} \Big) = A \otimes \overline{X} \otimes C$$

In view of Lemma 6 interval max-min matrix equation (4) is A-universally solvable. □

Theorem 8. *Interval max-min matrix equation* (4) *is universally solvable if and only if interval max-min matrix equation* (4) *is strongly solvable and*

$$\underline{A} \otimes \underline{X} \otimes \underline{C} = \overline{A} \otimes \overline{X} \otimes \overline{C}. \tag{13}$$

Proof. If interval max-min matrix equation (4) is strongly solvable and $\underline{A} \otimes \underline{X} \otimes \underline{C} = \overline{A} \otimes \overline{X} \otimes \overline{C} = B$ then $B = \underline{A} \otimes \underline{X} \otimes \underline{C} \leq A \otimes X \otimes C \leq \overline{A} \otimes \overline{X} \otimes \overline{C} = B$ for each $X \in \boldsymbol{X}$. We get $A \otimes X \otimes C = B \in \boldsymbol{B}$ for each $A \in \boldsymbol{A}$, for each $C \in \boldsymbol{C}$, and for each $X \in \boldsymbol{X}$, i. e., (4) is universally solvable.

The converse implication is trivial. □

Example 2. Let $\mathcal{I} = [0, 10]$ and let

$$\boldsymbol{A} = \begin{pmatrix} [1,3] & [5,8] \\ [1,2] & [4,6] \\ [4,7] & [2,3] \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} [3,5] & [2,5] \\ [3,5] & [4,5] \\ [3,6] & [2,6] \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} [4,6] & [6,7] \\ [3,3] & [3,4] \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} [2,4] & [3,3] \\ [4,4] & [3,5] \end{pmatrix}.$$

We check all above defined solvability concepts.

Solution:
First, we check the universal and strong solvabilities. We have

$$\underline{A} \otimes \underline{X} \otimes \underline{C} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \\ 3 & 3 \end{pmatrix} \geq \underline{B}, \quad \overline{A} \otimes \overline{X} \otimes \overline{C} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \\ 4 & 4 \end{pmatrix} \leq \overline{B}, \quad \underline{A} \otimes \underline{X} \otimes \underline{C} \neq \overline{A} \otimes \overline{X} \otimes \overline{C}.$$

According to Theorem 2 and Theorem 8 the given interval matrix equation is strongly solvable, but it is not universally solvable.

For checking the X-universal solvability we compute

$$X^{(11)} = \begin{pmatrix} 4 & 3 \\ 4 & 3 \end{pmatrix}, \quad X^{(12)} = \begin{pmatrix} 2 & 3 \\ 4 & 3 \end{pmatrix}, \quad X^{(21)} = \begin{pmatrix} 2 & 3 \\ 4 & 3 \end{pmatrix}, \quad X^{(22)} = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}.$$

We have

$$\underline{A} \otimes X^{(11)} \otimes \underline{C} = \overline{A} \otimes X^{(11)} \otimes \overline{C} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \\ 4 & 4 \end{pmatrix} \quad \text{and}$$

$$\underline{A} \otimes X^{(jl)} \otimes \underline{C} = \overline{A} \otimes X^{(jl)} \otimes \overline{C} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \\ 3 & 3 \end{pmatrix} \quad \text{for each } (j,l) \in \{(1,2),(2,1),(2,2)\}.$$

Since (10) is satisfied for each $j \in N$, $l \in S$, the given interval system is X-universally solvable.

At least, we show that the given interval system is not A-universally solvable. Let us take

$$A^{(11)} = \begin{pmatrix} 3 & 5 \\ 1 & 4 \\ 4 & 2 \end{pmatrix} \quad \text{and} \quad C^{(11)} = \begin{pmatrix} 6 & 6 \\ 3 & 3 \end{pmatrix}$$

Since

$$A^{(11)} \otimes \underline{X} \otimes C^{(11)} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \\ 3 & 3 \end{pmatrix} \quad \text{and} \quad A^{(11)} \otimes \overline{X} \otimes C^{(11)} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \\ 4 & 4 \end{pmatrix},$$

according to Theorem 7 the given interval system is not A-universally solvable. √

4.1 Conclusion

In this paper, we dealt with the solvability of interval matrix equations in max-min algebra. Max-min algebra is a useful tool for describing real situation in the economy and industry. In Example 1, the values a_{ij}, x_{jl}, and c_{lk} represent the capacities of corresponding connections. In economics, those values can represent for example the financial costs for the production or transporting of some products. In another example, a_{ij} represents a measure of the preference of the property P_i of some object before the property Q_j, similarly x_{jl} (c_{lk}) represent a measure of preference of property Q_j to property T_l (property T_l to property D_k). There are many possibilities for definition of more solvability concepts. Our main objective for the future will be studying another types of solvability of interval system with bounded and unbounded solution.

References

[1] Draženská, E., Myšková, H.: Interval fuzzy matrix equations. *Kybernetika* **53**, no. 1 (2017), 99–112.

[2] Myšková, H.: Interval systems of max-separable linear equations. *Lin. Algebra Appl.* **403** (2005), 263–272.

[3] Myšková, H: The algorithm for testing solvability of max-plus interval systems. In: *Proceedings of 29th International Conference on Mathematical Methods in Economics 2011* (M. Dlouhý and J. Skočdopolová, eds.), University of Economomics, Faculty of Informatics and Statistics, Jánska dolina, 2011, 498–503.

[4] Myšková, H.: Max-min interval systems of linear equations with bounded solution. *Kybernetika* **48**, no. 2 (2012), 299–308.

[5] Myšková, H.: Interval fuzzy matrix equations. In: *Proceeding of 34th International Conference Mathematical Methods in Economics*, (Kocourek, A., Vavrousek, M. eds.), Technical University of Liberec, Faculty of Economics, Liberec, 2016, 588–593.

[6] Nola, A. Di, Salvatore, S., Pedrycz, W., Sanchez, E.: *Fuzzy Relation Equations and Their Applications to Knowledge Engineering*. Kluwer Academic Publishers, Dordrecht, 1989.

[7] Plavka, J: On the $O(n^3)$ algorithm for checking the strong robustness of interval fuzzy matrices. *Discrete Appl. Math.* **160** (2012), 640–647.

[8] Sanchez, E: Medical diagnosis and composite relations. In: *Advances in Fuzzy Set Theory and Applications*, North-Holland, Amsterdam, The Netherlands, 1979, 437–444.

[9] T. Terano, Y. Tsukamoto: Failure diagnosis by using fuzzy logic. In: *Proc. IEEE Conference on Decision Control*, New Orleans, 1977, 1390–1395.

[10] Zimmermann, K.: *Extremální algebra*. Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV, Praha, 1976.

Assessment of the Impact of Traffic Police Preventive Interventions

Dana Nejedlová[1]

Abstract. Free publicly available datasets describing weather and traffic accidents in the Czech Republic have been used for the training of a feed-forward neural network so that it could predict the level of the number of traffic accidents and their cost from weather, weekday, and month. The neural network learns the data for each of the 14 Czech regions separately and also the idea of cross-validation is utilized. The data for each learning task have been separated into Training Set, Development Test Set, and Test Set. Then a statistically significant number of experiments with neural network to get the accuracy of the prediction of the Test Set that happens when the accuracy of the Development Test Set is maximized have been conducted. The aim of the research is to learn whether this methodology can statistically detect any significant difference of the accuracy of prediction between the Test Set formed from days with interventions reported by the Czech Police and the randomly selected Test Set using the assumption that the neural network learns dependencies not affected by preventive interventions.

Keywords: Cross-validation, statistical hypothesis testing, feed-forward neural network, traffic accidents.

JEL Classification: C32, C45, R41, R50
AMS Classification: 68T, 62H

1 Introduction

It seems to be obvious that preventive traffic police interventions reduce the number of traffic accidents – after all, it is their purpose to do so. This work attempts at scientific treatment of them. The research on this topic worldwide confirms their positive effects on the reduction of accidents. Found studies focus on particular elements increasing security like safety belts [4], road adjustments [2], legislation, educational campaigns [9], the effect of police interventions on crime mitigation in particular geographical areas [1], the effect of sobriety checkpoints in reducing alcohol-related fatalities and injuries [5] by means of systematic reviews of existing studies. An attempt for economic analysis of effectiveness of anti-alcohol interventions is given in study [3] using a methodology presented in [10]. Article [14], which compares the number of deaths caused by traffic accidents per population of 100,000 people before and after the enforcement of usage of seat belts, motorcycle helmets, general traffic laws, and mass media educational campaigns, is perhaps closest to the research presented in this paper. Both researches study a particular country and use its national traffic police data. While [14] studies the effect of long-term educational and law enforcement programs, this paper tries to assess the immediate local effect of traffic police interventions on the number of traffic accidents and related economic damage.

2 The aim of the research and its methodology

The aim of this research is to find out whether the use of neural network and statistical tests of its output could detect any significant difference between days with and without preventive police interventions in terms of the number of traffic accidents and related economic damage.

While statistical tests like Pearson's chi-squared tests [15] can detect the dependence of the pair of variables, e.g. temperature on the number of accidents, neural network can work with multiple variables. Feed-forward neural network reads a number of input variables and adjusts its parameters so that its output is as close as possible to some other variables that should be predicted. Statistically prevailing dependencies determine its working. When neural network learns the number of accidents on days among which only a small part are days with traffic police interventions, its prediction is not affected by traffic police interventions. This property of neural network can be used to compare the number of traffic accidents and related economic damage on days with and without preventive traffic police interventions. If the accuracy of its prediction on days with interventions is significantly lower than its accuracy on days without interventions, it could mean that the interventions made some difference.

[1] Technical University of Liberec, Faculty of Economics, Department of Informatics, Studentská 1402/2, 461 17 Liberec 1, Czech Republic, e-mail: dana.nejedlova@tul.cz.

3 Data

Datasets used in this study include weather from the National Centers for Environmental Information [6, 7], traffic accidents published by the Police of the Czech Republic on their web page [12], and reports on preventive traffic police interventions [13]. Because the immediate effect of traffic police interventions is local, all data should also be local. Available data sources allow division of data into 14 Czech regions, see e.g. [8].

3.1 Weather

Properties and processing of weather data are published in paper [11] which describes the research on which the research published in this paper is based on.

Weather dataset [6, 7] is structured by several Czech meteorological stations. For each of the 14 Czech regions the closest meteorological station has been used to get the data about mean temperature, precipitation and snow depth in this region for a particular day. Precipitation and snow depth have been transformed into binary values meaning either no or some precipitation or lying snow throughout a day in a particular region.

3.2 Traffic accidents

Of all available information in traffic accidents dataset [12] the total number of accidents and the total cost of economic damage per day for each of the 14 Czech regions have been selected as the values that will be predicted by the neural network.

Like in [11], the nominal values of the number of accidents and economic damage have been transformed into three categories "low", "medium", and "high" because this kind of output is the most appropriate for the neural network to learn. Before this transformation, like in [11], the nominal values in each of the 2 datasets had been fitted by quadratic curve and transformed into differences from this curve because the trend of these values has a rising tendency throughout the years. Unlike in [11], the trend curve has been estimated for each region separately using all available days including the Test Set. The Test Set has been included because in the current task we do not question the ability of the neural network to predict future events from past events but its ability to detect different accident incidence by the difference in its accuracy for 2 methods of forming the Test Set:

- Method 1, referred to as "Interventions" in further text, forms the Test Set from days with interventions.
- Method 2, referred to as "Random" in further text, forms the Test Set merely from randomly chosen days.

After getting values in the form of differences from trends, 2 border values for the number of accidents and 2 border values for economic damage dividing these values into 3 above-mentioned categories have been stated in each region separately, so that each region has approximately the same number of values in each category for each of these 2 datasets.

3.3 Preventive traffic police interventions

Reports on preventive traffic police interventions in [13] have to be read through to get a list of days on which preventive traffic police interventions happened. Of all preventive police activities, only the situations when traffic police went out on some road and interacted with drivers have been selected. Reading through police reports is very labor-intensive because of a large number of these reports and because of different structures of web presentations on different Czech regional police departments. Some regional police departments have their preventive activities separated from other reports and some departments have only a list of all reports which includes all police activities, not only prevention of the traffic police. Currently there are some 50 thousand reports that should yet be read through to get more days with preventive traffic police interventions on the list. See Table 1 for details.

3.4 Time span and division into Training, Development Test, and Test sets

Previous research [11] uses data from 2009.02.10 to 2016.04.19. Current research uses days until 2017.04.13. This means that for this research 2985 days can be used.

For each region 3 sets have been created. Training Set is formed from 2000 oldest days. Development Test Set is formed from 631 newer days. Test Set is formed from 354 days. If Method 1 (see Chapter 3.2) is used, all days with some preventive intervention in the region are present among these 354 days. The intention of this division is to have these three sets of the same size in each of the Czech regions so that the results of further processing are comparable. All days in the Test Set of the "Liberecký" region are days with preventive traffic police interventions. All other Czech regions have in their Test Set besides all found days with interventions some filler days chosen at random. See the number of found days with interventions in Table 1.

Region	Number of found days with interventions	Total number of reports	Number of reports not yet processed	Web
Jihočeský	283	917	0	http://www.policie.cz/krajske-reditelstvi-jihoceskeho-kraje-prevence.aspx
Jihomoravský	23	22174	21454	http://www.policie.cz/sprava-jihomoravskeho-kraje-zpravodajstvi.aspx
Karlovarský	71	256	0	http://www.policie.cz/akce-a-projekty-prevence.aspx http://www.policie.cz/akce-a-projekty-akce.aspx
Královéhradecký	280	23970	16006	http://www.policie.cz/kralovehradecky-kraj-zpravodajstvi.aspx http://www.policie.cz/clanek/krajska-reditelstvi-policie-kralovehradecky-kraj-archiv-zpravodajstvi-archiv-zpravodajstvi.aspx
Liberecký	354	14172	0	http://www.policie.cz/krajske-reditelstvi-policie-lbk-zpravodajstvi.aspx
Moravskoslezský	186	254	0	http://www.policie.cz/sprava-severomoravskeho-kraje-akce-a-projekty.aspx
Olomoucký	340	800	0	http://www.policie.cz/prevence-v-olomouckem-kraji.aspx http://www.policie.cz/zpravodajstvi-olk.aspx
Pardubický	101	865	0	http://www.policie.cz/clanek/akce-a-projekty-668803.aspx
Plzeňský	170	482	0	http://www.policie.cz/sprava-zapadoceskeho-kraje-akce-a-projekty.aspx
Praha	128	365	0	http://www.policie.cz/cinnost-a-akce.aspx
Středočeský	61	5836	4515	http://www.policie.cz/sprava-stredoceskeho-kraje-zpravodajstvi.aspx
Ústecký	80	715	0	http://www.policie.cz/sprava-severoceskeho-kraje-akce-a-projekty.aspx
Vysočina	88	340	0	http://www.policie.cz/krajske-reditelstvi-policie-kvs-prevence.aspx
Zlínský	12	8694	8279	http://www.policie.cz/clanek/zpravodajstvi-policie-zlinskeho-kraje-150757.aspx
Total	2177	79840	50254	List of found reports on author's page

Table 1 Data sources for traffic police interventions existing on 2017.05.03

4 Algorithm for the detection of the effect of preventive interventions

Neural network is used to predict the level of the number of traffic accidents and related economic damage for each day. As mentioned in Chapter 3.2, this level is encoded into 3 categories, and neural network learns to predict the correct category from the input data for a particular day. A process of learning of the single dataset will be referred to as "experiment" in further text. The neural network starts its training from the state with random weights which are its parameters. The result of its training depends on its initial state. It is possible to get the reliable information about the accuracy of its predictions only as a result of statistically significant number of experiments. That is why the neural network has been learning each region and dataset 100 times. In each experiment the neural network learns 3000 times each day in the Training Set and records the accuracy on the Test Set when its accuracy on the Development Test Set is maximal. The results of these experiments are statistically tested for significant difference between two methods of forming the Test Set explained in Chapter 3.2.

4.1 Neural network

A feed-forward neural network which has been programmed in C language for study [11] has been adjusted for this study. As in [11], it predicts separately the category of the number of accidents and the level of economic

damage from month, day of week, temperature 219 days ago, temperature 32 days ago, temperature for the current day, precipitation and snow depth using 3 sets of data, the Training Set, Development Test Set, and Test Set.

4.2 Statistical test

100 experiments with the Test Set formed by Method 1 "Interventions" and 100 experiments with the Test Set formed by Method 2 "Random" produce two sets of the number of correctly predicted categories. The Welch test (1) [15] compared to Student's *t*-distribution with *v* degrees of freedom (2) [15] has been used for testing whether their means (averages) are the same, which is the null hypothesis. This hypothesis is rejected if the *p*-value of the absolute value of the Welch test $|W|$ is smaller than probability 0.01, which means that the probability of rejecting this hypothesis when it is in fact true (i.e. the probability of Type I error) is not higher than 0.01. Symbols in equations (1), (2), and (3) with index 1 are for the results of Method 1 and symbols with index 2 are for the results of Method 2.

$$W = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{1}$$

The number of correctly classified categories will be used as *X* variables in (1). The other symbols in (1) are standard deviations *s*, i.e. variances are s^2, and *n* is the number of values, which is equal to 100.

The degrees of freedom looked up in Student's *t*-distribution table are given by equation (2) [15].

$$v = \frac{(q_1 + q_2)^2}{\frac{q_1^2}{n_1 - 1} + \frac{q_2^2}{n_2 - 1}} \tag{2}$$

Symbols q_1 and q_2 in (2) are given by relations (3) [15].

$$q_1 = \frac{s_1^2}{n_1} \text{ and } q_2 = \frac{s_2^2}{n_2} \tag{3}$$

P-values of Student's *t*-distribution table have been computed in MS Excel by function "TDIST(ABS(*W*),*v*,2)". In versions newer that Excel 2007 function "T.DIST.2T(ABS(*W*),*v*)" should be used.

5 Experimental results

Results of neural network training are shown in Tables 2 and 3.

Region	Interventions	Random	Welch test	Degrees of freedom	*p*-value
Jihočeský	102.89	105.05	-1.987550	180.9364	0.048377
Jihomoravský	120.71	123.75	-2.563290	194.1532	0.011126
Karlovarský	88.36	87.30	0.990696	197.9978	0.323049
Královéhradecký	**99.19**	**103.57**	**-3.876580**	**181.8908**	**0.000148**
Liberecký	**98.05**	**102.89**	**-4.890000**	**187.0085**	**0.000002**
Moravskoslezský	129.49	132.19	-2.480470	193.3255	0.013977
Olomoucký	110.21	108.80	1.325444	173.5115	0.186772
Pardubický	105.86	106.28	-0.343750	197.9976	0.731399
Plzeňský	100.72	98.53	1.929005	196.5972	0.055175
Praha	**206.17**	**216.16**	**-7.868450**	**188.6962**	**2.75E-13**
Středočeský	122.19	120.27	1.573078	193.5211	0.117339
Ústecký	**123.19**	**126.95**	**-3.072680**	**196.4378**	**0.002423**
Vysočina	99.73	99.76	-0.023600	196.6765	0.981195
Zlínský	107.05	107.30	-0.212290	197.7608	0.832097
Average	115.27	117.06	-1.535746	191.1794	0.235934

Table 2 Average number of correctly classified categories of the number of accidents in the Test Set formed by methods "Interventions" and "Random" in 100 experiments. When *p*-value is smaller than probability 0.01, the differences in the results of the 2 methods of training are considered statistically significant and their rows are displayed in bold.

Region	Interventions	Random	Welch test	Degrees of freedom	*p*-value
Jihočeský	99.77	97.83	1.508743	190.0103	0.133026
Jihomoravský	110.40	109.37	0.813440	195.8868	0.416958
Karlovarský	86.98	86.96	0.010070	175.3466	0.991977
Královéhradecký	**86.39**	**93.20**	**-5.422930**	**195.4249**	**1.72E-07**
Liberecký	**86.29**	**90.67**	**-3.502020**	**190.1538**	**0.000576**
Moravskoslezský	111.71	110.69	0.828421	193.7296	0.408456
Olomoucký	95.38	98.34	-2.363840	181.2100	0.019146
Pardubický	99.58	96.64	2.181468	197.8164	0.030333
Plzeňský	92.94	95.38	-1.971900	197.1722	0.050020
Praha	**164.06**	**168.78**	**-3.450240**	**195.8160**	**0.000686**
Středočeský	109.37	108.17	0.858778	196.4307	0.391512
Ústecký	113.31	112.62	0.500988	197.5765	0.616939
Vysočina	92.22	95.20	-2.515220	197.3088	0.012695
Zlínský	91.61	91.33	0.188017	194.4475	0.851060
Average	102.86	103.94	-0.881159	192.7379	0.280242

Table 3 Average number of correctly classified categories of the cost of economic damage in the Test Set formed by methods "Interventions" and "Random" in 100 experiments. When *p*-value is smaller than probability 0.01, the differences in the results of the 2 methods of training are considered statistically significant and their rows are displayed in bold.

We can say from Tables 2 and 3 that in regions where the *p*-value of the absolute value of the Welch test is smaller than probability 0.01 the traffic police preventive interventions have significantly influenced the number of accidents or economic damage with the probability of 1 – 0.01, i.e. with a 0.99 level of confidence.

6 Discussion of results

The average accuracy for all regions equal to the Average of Method 2 "Random" from the lower row of Tables 2 and 3 divided by the size of the Test Set equal to 354, as stated in Chapter 3.4, is equal to 33% for the number of accidents and to 29% for the economic damage. In study [11] the accuracy of prediction of the category of the number of accidents is in the range of 53% to 64% and the accuracy of prediction of the category of the cost of economic damage is in the range of 43% to 50% of days in the Test Set. Inspection of graphs with the values of the number of accidents and the values of the cost of economic damage with their quadratic trends suggests that the poor results of the neural network in this research is caused by bad fit of the quadratic trend to the data in some regions. Only the "Praha" region has the accuracy of all of its datasets within the range of the results in [11] and its quadratic trends seem to be fitting well both the number of accidents and the cost of economic damage datasets.

However, the poor accuracy of neural network may not hinder its ability to detect the effect of preventive interventions. The following evidence about the method of assessment of preventive interventions presented in this paper can be derived from Tables 1, 2, and 3.

Facts that speak against the methodology:

- In some regions the accuracy of "Random" is lower than the accuracy of "Interventions". The rationale why the accuracy of the recognition of the "Random" dataset should be higher than the accuracy of the recognition of the "Interventions" dataset is explained in Chapter 2.
- Relatively large quantities of days with interventions do not always result in recognition of "Random" being more accurate than the recognition of "Interventions" dataset, see e.g. the "Olomoucký" region in Tables 1 and 2.

Facts that speak in favor of the methodology or at least make some sense:

- The average number of correctly recognized categories for all regions equal to the Average from the lower row of Tables 2 and 3 is higher in the "Random" method than in the "Interventions" method.
- In all statistically significant regions (in bold) the number of correctly recognized categories of the "Random" method is higher than the number of correctly recognized categories of the "Interventions" method.
- Statistically significant regions in Table 3 are subset of statistically significant regions in Table 2. This makes sense because the recognition of economic damage has always been more inaccurate than the recognition of the number of accidents.

- The "Praha" region, that has the highest accuracy in both "Interventions" and "Random" methods and in both the number of accidents and the economic damage datasets, has the smallest area of the all Czech regions, see e.g. [8]. Most interventions did not cover the area of the whole region, so interventions in smaller region should have greater impact for that region.

7 Conclusion

The original method based on statistical processing of the results of the prediction using neural network for the detection of the impact of traffic police preventive interventions has been presented. According to this method, preventive interventions in the period of 2009.02.10 to 2017.04.13 in regions "Královéhradecký", "Liberecký", and "Praha" seem to have significant impact on the number of traffic accidents and the cost of economic damage. Possible improvements to this method include finding more days or more appropriate days with preventive interventions and higher order, maybe cubic, polynomial for regression of the time series of predicted variables.

References

[1] Braga, A. A.: Hot Spots Policing and Crime Prevention: A Systematic Review of Randomized Controlled Trials. *Journal of Experimental Criminology* **1** (2005), 317–342.

[2] Bunn, F., Collier, T., Frost, C., Ker, K., Roberts, I., and Wentz, R.: Traffic Calming for the Prevention of Road Traffic Injuries: Systematic Review and Meta-Analysis. *Injury Prevention* **9** (2003), 200–204. doi: 10.1136/ip.9.3.200.

[3] Cobiac, L., Vos, T., Doran, Ch., and Wallace, A.: Cost-Effectiveness of Interventions to Prevent Alcohol-Related Disease and Injury in Australia. *Addiction* **104** (2009), 1646–1655. doi: 10.1111/j.1360-0443.2009.02708.x.

[4] Dinh-Zarr, T. B., Sleet, D. A., Shults, R. A., Zaza, S., Elder, R. W., Nichols, J. L., Thompson, R. S., Sosin, D. M., and the Task Force on Community Preventive Services: Reviews of Evidence Regarding Interventions to Increase the Use of Safety Belts. *Am J Prev Med* **21**, 4S (2001), 48–65.

[5] Fell, J. C., Lacey, J. H., and Voas, R. B.: Sobriety Checkpoints: Evidence of Effectiveness Is Strong, but Use Is Limited. *Traffic Injury Prevention* **5** (2004), 220–227. doi: 10.1080/15389580490465247.

[6] Menne, M. J., Durre, I., Korzeniewski, B., McNeal, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R. S., Gleason, B. E., and Houston, T. G.: Global Historical Climatology Network - Daily (GHCN-Daily), Ver-sion 3. [FIPS:EZ - Czech Republic - 2000-01-01--2016-03-31]. NOAA National Climatic Data Center. (2012), doi: 10.7289/V5D21VHZ.

[7] Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G.: An Overview of the Global Historical Climatology Network-Daily Database. J. Atmos. *Oceanic Technol* **29** (2012), 897–910. doi: 10.1175/JTECH-D-11-00103.1.

[8] Ministry of Regional Development of the Czech Republic: Portal of Spatial Planning. (2017). Retrieved May 7, 2017, from http://portal.uur.cz/spravni-usporadani-cr-organy-uzemniho-planovani/kraje.asp.

[9] Morrison, D. S., Petticrew, M., and Thomson, H.: What are the Most Effective Ways of Improving Population Health through Transport Interventions? Evidence from Systematic Reviews. *J Epidemiol Community Health* **57** (2003), 327–333. doi: 10.1136/jech.57.5.327.

[10] Murray, Ch. J. L., Evans, D. B., Acharya, A., and Baltussen, R. M. P. M.: Development of WHO Guidelines on Generalized Cost-Effectiveness Analysis. *Health Economics* **9** (2000), 235–251. doi: 10.1080/15389580490465247.

[11] Nejedlová, D.: Prediction of Daily Traffic Accident Counts and Related Economic Damage in the Czech Republic. In: Aleš Kocourek, Miroslav Vavroušek (Editors) *MME 2016 – Proceedings of the 34th international conference Mathematical Methods in Economics.* 1st edition. Liberec 2016, 606–611.

[12] Police of the Czech Republic: Traffic Accidents Statistics. (2017). Retrieved April 19, 2017, from http://aplikace.policie.cz/statistiky-dopravnich-nehod/default.aspx.

[13] Police of the Czech Republic: Depart-ments with Territorially Defined Powers. (2017). Retrieved May 3, 2017, from http://www.policie.cz/clanek/utvary-s-uzemne-vymezenou-pusobnosti-794179.aspx

[14] Soori, H., Royanian M., Zali, A. R., and Movahedinejad, A.: Road Traffic Injuries in Iran: The Role of Interventions Implemented by Traffic Police. *Traffic Injury Prevention* **10** (2009), 375–378. doi: 10.1080/15389580902972579.

[15] Wilcox, R. R.: *Basic Statistics*. Oxford University Press, Inc., New York, 2009.

Modelling effective corporate tax rate in the Czech Republic

Daniel Němec[1], Silvester Dulák[2]

Abstract. Comparing the statutory corporate tax rates with the effective corporate tax rates at corporate level, one can observe a statistically significant average difference of 2 percent in the Czech Republic. The goal of our paper is to identify and quantify the most important factors influencing the effective corporate tax rate in the Czech Republic which contribute to the variability in the effective taxation. Using the public data from the accounting statements from the last 15 years we have formulated and estimated a panel data model using various real and financial indicators of the Czech companies. The most significant factors are the size of companies, the ratio of non-current assets to total assets, the rentability of assets, the ratio of inventory to the total assets, the economic activity classification CZ-NACE, and the legal form. We show that the influence of the indebtedness ratio is non-linear. Surprisingly, the company's age and the retained losses prior years have no effects on final effective tax rates. Our model is able to predict statutory tax rate pretty well.

Keywords: effective corporate tax rate, statutory corporate tax rate, panel data models, Czech Republic.

JEL classification: C33, H21, H25
AMS classification: 62P20, 91B38

1 Introduction

Tax burden is very often expressed by the tax revenue to the GDP (gross domestic product) ratio. This indicator is related to the effective taxation as well. It may differ from the statutory tax rate. But, such a kind of aggregation brings many problems when analysing analyse the factors influencing the effective tax rate (ETR)in the economy and the resulting tax incidence. Buijink et al. [2] compared the weighted statutory tax rates (STR) in European union (EU) countries with the median effective tax rates of these countries computed using the corporate level data set. The resulting differences (within the period 1990-1996) ranged between 1.07% for Sweden and 22.1% for Portugal (the average difference in EU was 9.59%). As for the Czech Republic, comparing the statutory corporate tax rates in the last 15 years with the effective corporate tax rates at corporate level, one can observe a statistically significant average difference of 2 percent.

The goal of our paper is to identify and quantify the most important factors influencing the effective corporate tax rate in the Czech Republic. Using the public data from the accounting statements from the period from 2004 to 2013, we have formulated and estimated a panel data models using various real and financial indicators of the Czech companies.

Statutory tax rate is applied on the tax base to get the final tax liability. Regarding the differences in the process of determining the tax base among the countries, there is a considerable problem related to plausibility of comparison of corporate tax rates of analysed countries. STR cannot be used for evaluating the real tax burden of companies. On the other hand, effective tax rate includes various aspects of tax legislation (e.g. tax allowances). As Fernandez-Rodriguez and Martinez-Arias [5] pointed out, the ETR and its development in comparison with the STR is crucial for improving the tax legislation. The methodology to compute the ETR is not unique. Schratzenstaller [9] defines two general methods to evaluate the effective tax rate: the calculation of fictitious measures (based on tax codes and tax rates) and the calculation of factual indicators (based on statistical data on tax payments). Although the final ETR is not unique, Giannini and Maggiu [6] stated, that "'the existence of different indicators is not, per se, a shortcoming of this kind of analysis, but simply reflects the fact that each indicator measures different things.'" Moreover, various measures may be less or more appropriate depending on the policy questions related to the ETR.

Our approach to quantify the factors influencing the ETR in the Czech Republic may be treated as the factual indicators method. It is based on publicly available data from balance sheets and profit and loss accounts. Our

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00, Brno, Czech Republic, nemecd@econ.muni.cz
[2]KPMG Slovakia, Dvořákovo nábrežie 10, 811 02 Bratislava, Slovakia
Masaryk University, Faculty of Economics and Administration, Department of Finance, Lipová 41a, 602 00, Brno, Czech Republic, 349038@mail.muni.cz

definition of firm level effective (corporate) tax rate is as follows:

$$\frac{\text{corporate income tax (corporate tax liability)}}{\text{EBIT}} \tag{1}$$

where $EBIT$ is defined as the earnings before interest and taxes. Corporate tax is regulated by the law no. 586/1992 Coll., legislation (setting the income tax for both natural and legal persons). In the 2013, corporate tax revenues (including social security contributions according to OECD methodology) constitutes 9% of all tax revenues in the Czech Republic. Without social security contributions, this share was 3.4% only. Although the corporate taxes play only a minor role in the tax revenues, their overall setting is important for the economic development.

Our contribution tries to reveal and quantify the factors standing behind the effective corporate tax rates in the Czech Republic. We are focusing on the role of the size of companies, on the effects of various financial indicators at corporate level, on the differences among various industries (based on the economic activity classification CZ-NACE that is compatible with the Eurostat [3] NACE classification), and on the impact of the legal form. There are many studies dealing with the modelling of the ERT in various countries at the corporate levels. Feeny et al. [4] modelled the effective taxation in Australia, Janssen [7] focused on the ETR in Netherland, and Noor et al. [8] quantified the factors of ETR of Malaysian public listed companies. There are no papers dealing with the multivariate (multifactor) modelling of effective corporate tax rate in the Czech Republic (at the corporate level). This paper tries to fill up this existing gap.

The structure of our paper is as follows. In the Section 2, we introduce the methodology and the data, including the discussion of variables used in other studies. Section 3 presents and discusses the results of our model. Section 4 concludes.

2 Methodology and data

Our approach for building and estimating the model of effective taxation is based on a standard panel data model approach. There are many variables which might be related to the effective tax rate. We are focusing on the variables that may be obtained from the balance sheets and profit and losses accounts. To get the final income tax, the merchandise revenues are modified into the net income before taxes by the items modifying the income (tax-deductible expenses), by the items deductible from the tax base (e.g. retained losses prior years, expenses on research and development) and by the tax abatements (e.g. tax abatement for employee with a disability). All these factors are less or more included in the final model. To be more specific, we try to include (among other variables) the following set of variables:

- Tax write-offs approximated by the book depreciation, the size of non-current assets or by the ratio of a part of assets to the total assets (all these factors are in general significant).
- Interests paid related to and approximated by the overall or specific indebtedness (the effects on ETR is ambiguous).
- Employee benefits approximated by the number of employees (this factor is not usually included in the models but it may be an important factor in the Czech Republic).
- Retained losses prior years that may be related directly or indirectly through the negative numbers in profit and loss statement.
- Other factors that may be difficult to get from the publicly available balance sheets (on the other hand tax-deductible reserves may be related indirectly to the total amount of reserves, or tax-deductible allowances may be connected indirectly with the total amount of receivables).

Equation 2 shows the structural form of our model. The variables are described in the table 1.

$$\begin{aligned} ETR_{it} = {} & \alpha_i + \beta_1 AGE_{it} + \beta_2 SIZE_{it} + \beta_3 SIZE_{it}^2 + \beta_4 CAPINT_{it} + \beta_5 CAPINT_{it}^2 \\ & + \beta_6 INVINT_{it} + \beta_7 LEV_{it} + \beta_8 LEV_{it}^2 + \beta_9 TAXLOSS_{it} + \beta_{10} ROA_{it} \\ & + \beta_{11\ldots28} SECTOR\_?_{it} + \beta_{29\ldots39} FORM\_?_{it} + \beta_{40\ldots42} EMPL\_?_{it} + \epsilon_{it} \end{aligned} \tag{2}$$

The size of firms (expressed by the amount of assets) may have ambiguous effects on the ETR. The effect of indebtedness is usually negative, but there might be some non-linearities in its relation to the ETR. Tax write-offs approximated are approximated by the capital intensity. Negative effects may be found in the literature, although some non-linearities may prevail. Economic activity classification CZ-NACE (accordign to Eurostat [3]) is represented by the corresponding dummy variables (see Table 2 for the meaning of particular categories). As for the legal form, we use the dummy variables for join-stock company ($FORM\_AS$), state-owned or public companies ($FORM\_SP$), for companies with foreign ownership ($FORM\_FOREIGN$), and the dummy variables for any other legal forms ($FORM\_OTHER$). To analyse the effect of number of employees (considered as a proxy to overall social benefits), we use four dummy variables. The smallest companies ($EMPL\_MICRO$) are those

Variable	Description	Expected effect	Definition
ETR	effective tax rate	NA	$\frac{\text{corporate income tax}}{\text{EBIT}}$
AGE	age of the company	-	number of years from establishing
SIZE	size of the company	?	logarithm of assets
CAPINT	capital intensity	-	$\frac{\text{non-current assets}}{\text{assets}}$
INVINT	inventory to toal assets ratio	0	$\frac{\text{inventory}}{\text{assets}}$
LEV	indebtedness	-	$\frac{\text{liabilities (external resources)}}{\text{assets}}$
TAXLOSS	retained tax losses prior years	-	
ROA	rentability of assets	?	$\frac{\text{EBIT}}{\text{assets}}$
SECTOR	economic activity classification CZ-NACE	?	sectoral dummies, basic category = A (agriculture, forestry and fishing)
YEAR	period (year)	+ or 0	time dummies, basic category = 2010
FORM	legal form of the company	?/-	legal form dummies, basic category = SRO (limited liability company)
EMPL	number of employees	-	categorical dummies, basic category = MICRO (less than 10 employees)

Table 1 Definition of the model variables

with less than 10 employees. Small companies ($EMPL\_SMALL$) have the number of employees ranging from 10 to 49 employees, medium companies ($EMPL\_MEDIUM$) are those with more than 50, but less than 249 employees, and the largest companies ($EMPL\_BIG$) have more than 249 employees.

All indicators were computed using the panel data from balance sheets and profit and loss accounts provided by the database Albertina (Bisnode) [1]. This database covers the data from business register of the Czech companies. We have used the full data set covering the period from 2004 to 2013. The original data set contains 177 201 unique companies. Data were filtered using the following criterions:

- Total assets equal total liabilities and equity (to eliminate possible errors in published account statements).
- EBIT is nonzero and the sum of assets is a positive number.
- Excluding companies with missing values of financial indicators and companies where indicators are with obvious errors.
- Including companies with positive (nonzero) effective tax rate where the effective tax rate is not greater than 1.

We use the companies with the balance sheets available for the periods of more than one year. The final restricted sample consists of 141 724 observations of 29 733 companies. In the year 2013, there were almost 370 000 companies (in the legal form of join-stock company or limited liability company). Our sample represents almost 10% ot these companies. It should be noted, that this sample represents the most important companies in the Czech republic regarding the amount of corporate taxes paid by them. All the models are estimated by the means of fixed effect estimator and random effect estimator using the software SAS. The validity of random effect model was evaluated using the standard Hausmann test of exogeneity. Due to limited applicability of the panel model tests of heteroskedasticity and autocorrelation (which are strongly restricted to a specific form of the problem tested), we use the robust standard errors for evaluating the variability of estimated parameters.

3 Effective taxation in the Czech Republic

Table 3 contains the model estimates using the full sample. The effect of time-invariant factors cannot be estimated using the fixed-effect model. As the estimates of the random effect model show, the results do not differ from those obtained from the fixed effect model too much. The value and the signs of estimated coefficients are very similar. The most obvious differences are related to the size of company and capital intensity. The fixed effects model suggests that both variables influenced negatively the effective tax rate. On the other hand, the random effects model allows to some extent non-linearities in their influence. To be more specific, the expected negative effect of firm size is negative at the total amount of asset exceeding 50 million of Czech crowns. As for the capital intensity, only the firms with the ratio of non-current assets to the total assets greater than 2.5 could exhibit positive influence of capital intensity on effective tax rate. But, the capital intensity cannot exceed the value of 1.

Category	Description
A	Agriculture, forestry and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam and air conditioning supply
E	Water supply: sewerage, waste management and remediation activities
F	Construction
G	Wholesale and retail trade, repair of motor vehicles and motorcycles
H	Transportation and storage
I	Accommodation and food service activities
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional scientific and technical activities
N	Administrative and support service activities
O	Public administration and defece, compulsory social security
P	Education
Q	Human health and social work activities
R	Arts, entertainment and recreation
S	Other service activities

Table 2 NACE codes as defined by the Eurostat [3]

The estimated effects are mostly in accordance with the theory. Economic activity matters. The positive signs (excluding three economic activities) suggests, that the companies in agricultural sector tends to have the lowest effective tax rate. On the other, the firms in the educational sectors have effective tax rate mostly above the average level. To be more specific, the difference between agricultural sector and the sector of education is 6.08% in average. Random effect model estimates highlighted the role of legal form. The limited liability companies tend to have ETR by 0.7% lower than join-stock companies, and by 3.1% lower than state-owned companies. Corporate effective tax rate is by 4.61% higher for the companies with foreign ownership (compared to the limited liability companies). This is a surprising result. It suggests, that the companies with foreign ownership are not more successful in tax optimization than the domestic limited liability companies. It might indicate, that the foreign owners are willing to tax their incomes from abroad in the Czech Republic due to favourable tax conditions. The age of companies does not contribute to the lower effective tax rate. It is rather insignificant. The negative effect of the company size was proved without any non-linearities in case of fixed effect model. The random effect model suggests the negative effect of the company size for the firms with the total assets exceeding 50 million of Czech crowns.

The companies with higher capital intensity have overall lower ERT although this marginal effect seems to be diminishing (as the random effect model suggests). The role of inventory was surprisingly significant. There are no other studies that were able to prove negative relationship between the amount of inventory and the ETR. The marginal effect of indebtedness on ETR is negative and decreasing (in absolute terms).

The time dummies have expected positive signs that is in accordance with the lowest statutory tax rate from the year 2010. Moreover, the value of estimated coefficients for the years 2005-2009 is decreasing. This decrease fully corresponds to the changes in statutory tax rates in this period. In case of years 2011 and 2012, there are no significant time effects. These results are in accordance with the fact, that the statutory tax rates in 2011 a 2012 were at the level of the base year 2010. The overall coefficient of determination was 40% (based on the fixed effect model), although it is hardly relevant in context of panel data models.

Table 4 shows the tests of reliability of our model in a way aimed at the comparison of statutory and expected effective tax rate with an absence of any other factors. We use the approach proposed by the Feeny et al. [4]. We have thus compared the model implied effective tax rates with the statutory tax rates in the period 2005-2009. The equivalence of both tax rates was proved in the years 2005, 2007 and 2009. But, in the years 2006 and 2008 the statistically significant difference lies below the value of 1%.

To check the robustness of the results, we have carried out the estimates for individual economic activities and for all years separately. Although the signs of all effects remained the same, we were able to observe significant

Variable	FE model	RE model	Variable	FE model	RE model
AGE		-0.0002***	$Y2005$	0.0687***	0.0697***
		(0.0001)		(0.0014)	(0.001)
$SIZE$	-0.0068***	0.013***	$Y2006$	0.0556***	0.0558***
	(0.0011)	(0.002)		(0.0012)	(0.0009)
$SIZE^2$		-0.0006***	$Y2007$	0.0504***	0.0497***
		(0.001)		(0.001)	(0.0008)
$CAPINT$	-0.0632***	-0.0792***	$Y2008$	0.0278***	0.0268***
	(0.0036)	(0.0048)		(0.0009)	(0.0008)
$CAPINT^2$		0.016***	$Y2009$	0.092***	0.0097***
		(0.0062)		(0.0008)	(0.0008)
$INVINT$	-0.0141***	-0.0309***	$Y2011$		
	(0.0046)	(0.0022)			
LEV	-0.0342***	-0.0361***	$Y2012$		
	(0.0032)	(0.0015)			
LEV^2	0.0014***	0.002***	$FORM\_AS$	NA	0.007***
	(0.003)	(0.003)		(NA)	(0.0013)
$TAXLOSS$			$FORM\_FOREIGN$	NA	0.0461***
				(NA)	(0.0102)
ROA	-0.1054**	-0.081***	$FORM\_SP$	NA	0.031**
	(0.0057)	(0.0024)		(NA)	(0.0135)
$SECTOR\_B$	NA	0.0498***	$FORM\_OTHER$	NA	-0.0132***
	(NA)	(0.0081)		(NA)	(0.0029)
$SECTOR\_C$	NA	0.0178***	$EMPL\_SMALL$	0.0026**	
	(NA)	(0.0019)		(0.0012)	
$SECTOR\_D$	NA	0.0215***	$EMPL\_MEDIUM$	0.0047**	
	(NA)	(0.0053)		(0.0019)	
$SECTOR\_E$	NA	0.0296***	$EMPL\_BIG$		0.0042*
	(NA)	(0.004)			(0.0023)
$SECTOR\_F$	NA	0.0236***	$SECTOR\_G$	NA	0.0262***
	(NA)	(0.0021)		(NA)	(0.0019)
$SECTOR\_H$	NA	0.0221***	$SECTOR\_I$	NA	0.017***
	(NA)	(0.0031)		(NA)	(0.0036)
$SECTOR\_J$	NA	0.0378***	$SECTOR\_K$	NA	
	(NA)	(0.003)		(NA)	
$SECTOR\_L$	NA	0.0224***	$SECTOR\_M$	NA	0.0363***
	(NA)	(0.0031)		(NA)	(0.0023)
$SECTOR\_N$	NA		$SECTOR\_O$	NA	
	(NA)			(NA)	
$SECTOR\_P$	NA	0.0608***	$SECTOR\_Q$	NA	0.0392***
	(NA)	(0.0067)		(NA)	(0.0042)
$SECTOR\_R$	NA		$SECTOR\_S$	NA	0.0379***
	(NA)			(NA)	(0.0058)

Standard deviation in parenthesis. *, **, *** means that the coefficients are statistically significant at 10%, 5%, and 1% level of significance respectively.

NA means that the variables cannot be included in fixed effect model. The variables without any values were not statistically significant at 10% level of significance.

Table 3 Model estimates (full sample)

Variable	Estimated effect	Expected effect	p-value
$Y2005$	0.0687 (0.0014)	0.07	0.3649
$Y2006$	0.0556 (0.0012)	0.05	0.0000
$Y2007$	0.0504 (0.0010)	0.05	0.6891
$Y2008$	0.0278 (0.0009)	0.02	0.0000
$Y2009$	0.0092 (0.0008)	0.01	0.3552

p-values correspond to the standard t-test of a population mean of the parameter.

Table 4 Model consistency with the statutory tax rate

differences in the levels of marginal effects among the sectors, especially in case of capital intensity, indebtedness, and ROA.

4 Conclusion

Our model was able to estimate ETR at the level of STR in most of the analysed periods. The most significant factors influencing the ETR in the Czech Republic are the negative effects of the size of company and the capital intensity (both results are in accordance with other studies). The rentability of assets was mostly negative. We have observed a negative impact of the ratio of inventory to the total assets. This result was not observed in other studies on this topics. The economic activity classification CZ-NACE, and the legal form played an important role in determining the ETR as well. We have proved that the influence of the indebtedness ratio is rather non-linear which is in accordance with other studies except the estimated negative sign. Surprisingly, the company's age and the retained losses prior years have no effects on final ETR.

Acknowledgements

This work was supported by the grant No. GA17-26705S of the Czech Science Foundation. This support is gratefully acknowledged.

References

[1] Albertina (Bisnode) database. www.albertina.cz.

[2] Buijink, W., Janssen, B., Schols, Y.: *Corporate Effective Tax Rates in the European Union*. MARC research report, Maastricht, MARC research report, April 1999.

[3] Eurostat: *NACE Rev. 2 – Statistical classification of economic activities in the European Community*. Office for Official Publications of the European Communities, Luxembourg, 2008.

[4] Feeny, S., Gillman, M., Harris, M. N.: Econometric Accounting of the Australian Corporate Tax Rates: a Firm Panel Example. *Cardif Economics Working Papers* (2005), E2005/16.

[5] Fernandez-Rodriguez, E., Martinez-Arias, M.: Determinants of Effective Tax Rate: Evidence for USA and the EU, *INTERTAX* **39** 8/9 (2011), 381–395.

[6] Giannini, S., Maggiulli, C.: Effective Tax Rates in the EU Commission Study on Company Taxation: Methodological Aspects, Main Results and Policy Implications, *Ifo Studien* **48** 4 (2002), 633–654.

[7] Janssen, B.: Corporate Effective Tax Rates in the Netherlands, *De Economist* **153** 1 (2005), 47–66.

[8] Noor, R., Mastuki, N., Bardai, B.: Corporate Effective Tax Rates: A Study On Malaysian Public Listed Companies. *Malaysian Accounting Review* **7** 1 (2008), 1–18.

[9] Schratzenstaller, M.: Effective Company Taxation in Poland – Some Methodological Considerations and Empirical Results, *Intereconomics: Review of European Economic Policy* **40** 2 (2005), 89-99.

Optimal Inserting Depot Visits into Daily Bus Schedules

Stanislav Palúch[1], Tomáš Majer[2]

Abstract. Bus scheduling problem is to arrange a given set of trips into bus schedules – sequences of trips performable by one bus. Two main objectives are minimization of used vehicles and minimization of dead kilometers. Slovak and Czech bus providers often require that vehicles visit depot during working day in order to deliver money or to make vehicle maintenance or refueling. Just mentioned procedure is called a roundabout route to depot.
Our last paper [5] proposed a method how to manage that exactly as many roundabout routes were inserted into bus schedule how many running boards this bus schedule contained. Proposed method used a maximum group matching formulation. A drawback of suggested procedure was that sometimes more than one roundabout route to depot was assigned to one bus schedule. Elimination of such cases required a manual intervention.
This paper designs a new exact method based on several repeated group matching formulations, every repetition extended with additional constraint prohibiting several depot visits in one bus schedule. Some computer experiments with Gurobi solver for corresponding MILP model are discussed.

Keywords: vehicle scheduling, mixed linear programming, group matching

JEL classification: C44
AMS classification: 90C15

1 Introduction

A bus trip in a regular regional passenger transport is defined by a quadruple (d_k, a_k, u_k, v_k) where d_k is departure time, a_k is arrival time, u_k is departure bus stop and v_k is arrival bus stop of the trip k.

Regional bus transport is going on a transportation network containing crossings, bus stops and direct streets between them. Transportation network can be modeled by an arc weighted digraph. Corresponding distance matrices containing time and spatial distances between every pair of bus stops can be calculated using e.g. Floyd algorithm.

Let $\mathbf{M} = \big\{m(u,v)\big\}$ be a time distance matrix determining the travel time $m(u,v)$ from bus stop u to bus stop v.

Trip j can be carried immediately after trip i by the same bus if

$$d_j \geq a_i + m(v_i, u_j), \tag{1}$$

i.e. if a bus after arriving to the arrival bus stop v_i of trip i can pull to the departure bus stop u_j of the trip j sufficiently early. In this case we will say, that the trip j can be linked after trip i and we will write $i \prec j$.

A running board of a bus is sequence of trips $i_1, i_2, \ldots, i_r$ such that for every k, $1 \leq k < r$ it holds $i_k \prec i_{k+1}$. By other words, a running board is a sequence of trips which can be carried by the same bus in one day. It represents a day schedule of work for one bus.

The goal of bus scheduling with minimum number of buses is the following: Given the set $\mathcal{S}$ of trips to arrange all trips from $\mathcal{S}$ into minimum number of running boards. Resulting set of running boards is called a bus schedule.

Relation $\prec$ on the set $\mathcal{S}$ can be modeled by a digraph $G = (\mathcal{S}, E)$ where $E = \{(i,j) | i \in \mathcal{S}, j \in \mathcal{S}, i \prec j\}$.

Suppose that $\mathcal{S} = \{1, 2, \ldots, n\}$. Denote by x_{ij} a decision binary variable with the following meaning

$$x_{ij} = \begin{cases} 1 & \text{if the trip } j \text{ is linked immediately after trip } i \text{ in a running board} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

[1]University of Žilina, Department of mathematical methods and operations analysis, Univerzitná 8215/1, 010 26 Žilina, e-mail: stanislav.paluch@fri.uniza.sk
[2]University of Žilina, Department of mathematical methods and operations analysis, Univerzitná 8215/1, 010 26 Žilina, e-mail: tomas.majer@fri.uniza.sk

Imagine that we have a bus schedule with $n = |\mathcal{S}|$ buses – every bus makes only one trip. This situation corresponds to $x_{ij} = 0$ for all $(i,j) \in E$. When we realized one linkage for $(i,j) \in E$ (what is indicated by setting $x_{ij} = 1$) we have saved one bus. Therefore the more variables are equal to 1 the less vehicles are used in corresponding bus schedule. In order to obtain a feasible bus schedule, at most one trip j can be linked after arbitrary trip i and at most one trip i can be linked before arbitrary trip j. Just mentioned facts lead to the following mathematical model.

$$\text{Maximize} \quad \sum_{\substack{ij,\\(i,j)\in E}} x_{ij} \tag{3}$$

$$\text{subject to:} \quad \sum_{\substack{i,\\(i,j)\in E}} x_{ij} \leq 1 \quad \text{for } j = 1,2,\ldots,n \tag{4}$$

$$\sum_{\substack{j,\\(i,j)\in E}} x_{ij} \leq 1 \quad \text{for } i = 1,2,\ldots,n \tag{5}$$

$$x_{ij} \in \{0,1\} \tag{6}$$

Mathematical model (3) – (6) is an instance of assignment problem, therefore condition (6) can be replaced by

$$x_{ij} \geq 0 \tag{7}$$

and linear program (3) – (5), (7) has still integer solution.

If $\{x_{ij}\}_{(i,j)\in E}$ is the optimum solution of (3) – (6) then it holds for the minimum number of buses r

$$r = |S| - \sum_{(i,j)\in E} x_{ij} \tag{8}$$

2 Model with depot visiting

Slovak and Czech bus providers often require that vehicles visit depot during working day in order to deliver money or to make a vehicle maintenance or refueling.

A visit to the depot can be modeled as a special trip with both arrival place and departure place equal to depot and arrival and departure times determined by time interval necessary for all procedures required. In most cases this time interval is at least 30 minutes long and can be considered as a safety break for a driver.

In practice, a visit to depot has no fixed time position. It is convenient to schedule it into weak traffic hours it order to avoid the possibility of raising the number of vehicles.

Suppose, we have calculated a bus schedule with the minimum number of vehicles r till now without roundabout routes to depot.

We propose following procedure in order to manage desired visits of all r buses in depot:

Construct the set D of dummy trips representing bus stays in depot. Define digraph

$$\overline{G} = (S \cup D, A) \tag{9}$$

where $S \cup D$ is the union of the set of original set of trips together with the set of all dummy trips – visits to depot. The arc set A of $\overline{G}$ is the set of all ordered pairs (i,j) of elements of $S \cup D$ such that $i \prec j$.

The goal is to arrange all trips from S and r trips from D into minimum number of running boards. We introduce two decision variables x_{ij} and z_k with following meaning:

$$x_{ij} = \begin{cases} 1 & \text{if the trip } j \in S \cup D \text{ is linked immediately after trip } i \in S \cup D \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$z_i = \begin{cases} 0 & \text{if the dummy trip } i \in D \text{ is chosen into a running board} \\ 1 & \text{if the dummy trip } i \in D \text{ is not used in any running board} \end{cases} \tag{11}$$

The number of chosen depot stays should be equal exactly to r (the minimum number of running boards without roundabout routes) therefore $|D| - \sum_{i\in D} z_i = r$

If the dummy trip $i \in D$ was not used then all variables x_{ij} have to be equal to zero for all $j \in S$. This is ensured by the constraint

$$z_i + \sum_{\substack{j,\\(i,j)\in A}} x_{ij} \leq 1 \tag{12}$$

since $z_i = 1$ for any unused trip $i \in D$.

Similarly if dummy trip $j \in D$ is not used in any running board then zero values of all x_{ij} for all $i \in S$ are guaranteed by

$$z_j + \sum_{\substack{i,\\(i,j)\in A}} x_{ij} \leq 1 \tag{13}$$

On the other hand, if dummy trip $i \in D$ is used what is indicated by $z_i = 0$ then the constraint (12) guarantees that at most for one $j \in S$ is $x_{ij} = 1$. Similarly, if trip $j \in D$ is used what is indicated by $z_j = 0$ then the constraint (13) guarantees that at most for one $i \in S$ is $x_{ij} = 1$.

Hence, now we deal with the following optimization problem:

$$\text{Maximize} \quad \sum_{\substack{ij\\(i,j)\in A}} x_{ij} \tag{14}$$

$$\text{subject to:} \quad \sum_{\substack{i,\\(i,j)\in A}} x_{ij} \leq 1 \quad \text{for } j \in S \tag{15}$$

$$z_j + \sum_{\substack{i,\\(i,j)\in A}} x_{ij} \leq 1 \quad \text{for } j \in D \tag{16}$$

$$\sum_{\substack{j,\\(i,j)\in A}} x_{ij} \leq 1 \quad \text{for } i \in S \tag{17}$$

$$z_i + \sum_{\substack{j,\\(i,j)\in A}} x_{ij} \leq 1 \quad \text{for } i \in D \tag{18}$$

$$|D| - \sum_{i\in D} z_i = r \tag{19}$$

$$x_{ij} \in \{0,1\} \quad \text{for all } i,j \text{ such that } (i,j) \in A \tag{20}$$

$$z_i \in \{0,1\} \quad \text{for all } i \in D \tag{21}$$

2.1 Design of the set D of dummy trips

A deficit function $f(x)$ is a function with domain $\langle 0, 1440)$ defined as follows

$$f(x) = |\{\, i \mid x \in \langle d_i, a_i \rangle, i \in S\}\,| \tag{22}$$

Every $x \in \langle 0, 1440)$ represents time in a day in minutes, value $f(x)$ is the number of trips running in time moment x. A deficit function can help us to determine rush and weak traffic time interval. Typical deficit function has its rush hours in the morning and after noon and weak hours between 10:00 and 12:00 o'clock before noon. Roundabout routes should be placed into weak hour in order not to raise the number of vehicles.

Remember that a trip is a quadruple (d_k, a_k, u_k, v_k) where d_k is departure time, a_k is arrival time, u_k is departure bus stop and v_k is arrival bus stop of the trip k.

Every dummy trip representing stay in depot should have departure and arrival place equal to depot. Let $W \subset S$ be a set of trips running in week hours. If $(d_i, a_i, depot, depot)$ is a dummy trip then its latest time position is such that there exist a trip (d_j, a_j, u_j, v_j) such that

$$a_i = d_j - m(depot, u_j) \tag{23}$$

r otherwise this dummy trip could be shifted later. Therefore the set D of possible dummy trips can be defined as follows:

$$D = \{\, (a - T_d, a, depot, depot) \quad | \quad a = d_j - m(depot, u_j), j \in W\}\,, \tag{24}$$

where T_d is time of stay in depot.

When creating digraph $\overline{G} = (S \cup D, A)$, exclude all arcs (i, j) such that $i \in D$ and $j \in D$ from arc set A in order to get away of possibility that two roundabout routes will be assigned to the same vehicle.

3 Prohibiting two or more roundabout routes in one running board

Let us summarize till now designed procedure. First we have calculated the minimum number r of buses using mathematical model (3) – (6) and equation (8). Then we have created a lot of dummy trips each of them representing a visit to depot. As the last step was mathematical model (14) – (21) which has chosen exactly r dummy trips without raising the number of buses. Unfortunately this procedure assigned more depot visits to a bus schedule while another bus schedule was left without any depot visit.

Let $i_1, i_2, \ldots, i_p, i_{p+1}, \ldots, i_q, \ldots, i_k$ is a running board containing two depot visits – namely dummy trips i_p, i_q. This running board correspond to variables

$$x_{i_1,i_2} = 1, x_{i_2,i_3} = 1, \ldots, \underbrace{x_{i_p,i_{p+1}} = 1, x_{i_{p+1},i_{p+2}} = 1, \ldots, x_{i_{q-1},i_q} = 1}_{\text{infeasible part containing two dummy trips}}, \ldots, x_{i_{k-1},i_k} = 1 \tag{25}$$

One way how to prohibit such infeasible part of running board is the following condition

$$x_{i_p,i_{p+1}} + x_{i_{p+1},i_{p+2}} + \cdots + x_{i_{q-1},i_q} \leq q - p - 2 \tag{26}$$

Number of summands in (26) is $q - p - 1$. Condition (26) guarantees that at least one variable of the left hand side of (26) will take the value 0 and therefore infeasible part of (25) will never occure.

Therefore our new exact procedure how to ensure visits to depot for all running boards is as follows:

Step 1.Calculate the minimum number r of buses using mathematical model (3) – (6) and equation (8).

Step 2.Choose the set $W \subseteq S$ running in weak hours. Create the set D of dummy trips representing depot visits using (24).

Step 3.Using model (14) – (21) calculate bus schedule with exactly r visits to depot.

Step 4.If every running board contains exactly one depot visit, STOP. Otherwise continue with Step 5.

Step 5.Identify running boards with multiple visits to depot. Add constraint (26) for every infeasible sequence of trips to model (14) – (21).

Step 6.Solve model (14) – (21) with all additional constraints. If number of running boards rises (value of criteria function decreases) STOP – there is no possibility to include required depot visits into bus schedule with r buses. It is necessary to add a bus. Otherwise go to Step 4.

4 Computational experiment

We used Gurobi MILP solver on Intel Xeon with 4 CPU Cores to solve linear model (14)–(21).

We used the same real world data for municipal bus public transport in Slovak town Martin–Vrútky as in our previous paper [5]. The public transport system of this town consisted of 725 trips organized in 18 lines. Corresponding deficit function is presented on Figure 1.

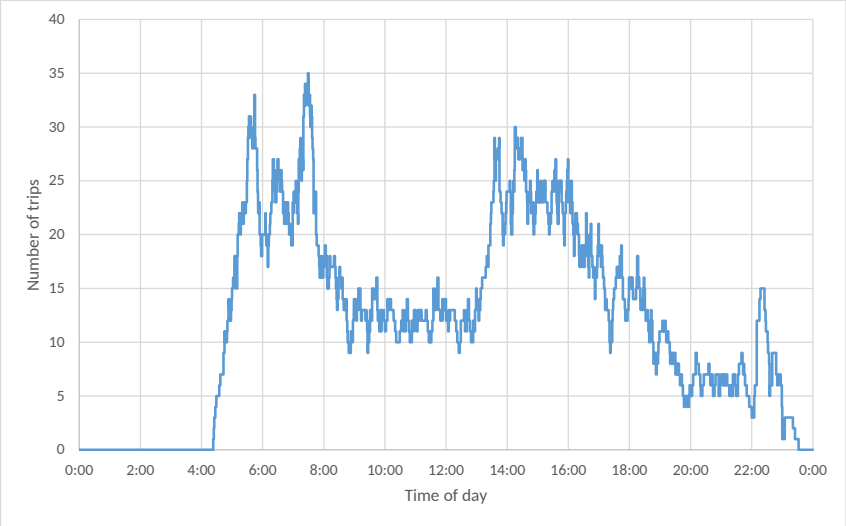


Figure 1 Deficit function of typical working day

Depot was identified with its closest bus stop 59 since input data did not contain an individual vertex for depot.

We set the time needed for operations in depot $T_d = 60$min and we used the deficit function for determining time window in which it is reasonable to define time positions of roundabout dummy trips. Time window between 8:00 and 12:00 – in hours with weak traffic was chosen. We have created 130 possible dummy trips.

First calculation of the minimum number of running boards (without roundabout routes to depot) solved by model (3)–(7) took approx. 56 seconds and yielded 39 buses. Therefore we had to use exactly $r = 39$ dummy trips in order to assign exactly one dummy trip to every running board in equation (19).

The result was not satisfactory since sometimes two dummy trips were placed into one running board and so another running boards remained without roundabout trip. Therefore we have identified running boards with more than one dummy trip and we have appended condition (26) prohibiting every multiple appearance of dummy trips in one running board to model (14) – (21). The result of extended model was not satisfactory therefore we added further additional constraints.

We repeated this procedure until every running board contained exactly one dummy trip. A feasible solution was obtained after 51 iterations and approx. 8.28 minutes. Full computational steps are shown on Table 1.

Iteration	1	2	3	4	5	6	7	8	9	10	11
Time of model constr. [s]	87.11	0.48	0.47	0.47	0.48	0.47	0.48	0.47	0.48	0.47	0.48
Time of model calc. [s]	9.35	10.21	10.18	9.93	8.35	10.03	8.87	8.11	8.88	10.02	9.87
Nr. of brds. without dep.	4	4	7	8	5	5	6	8	5	5	4
Iteration	12	13	14	15	16	17	18	19	20	21	22
Time of model constr. [s]	0.47	0.47	0.48	0.48	0.47	0.47	0.48	0.48	0.48	0.47	0.48
Time of model calc. [s]	9.38	8.97	8.82	10.29	10.21	9.64	9.71	9.05	8.76	7.97	9.33
Nr. of brds. without dep.	4	4	4	3	4	3	3	3	7	6	6
Iteration	23	24	25	26	27	28	29	30	31	32	33
Time of model constr. [s]	0.47	0.48	0.49	0.50	0.47	0.48	0.47	0.47	0.47	0.47	0.48
Time of model calc. [s]	9.57	9.40	10.08	9.28	10.07	8.41	9.59	9.78	9.56	9.17	10.72
Nr. of brds. without dep.	4	4	3	2	5	6	5	4	4	2	4
Iteration	34	35	36	37	38	39	40	41	42	43	44
Time of model constr. [s]	0.47	0.47	0.47	0.47	0.50	0.47	0.47	0.47	0.47	0.47	0.47
Time of model calc. [s]	10.67	9.12	9.74	10.36	9.87	9.46	10.06	8.92	8.25	8.64	8.86
Nr. of brds. without dep.	5	6	7	6	4	4	4	5	4	3	2
Iteration	45	46	47	48	49	50	51				
Time of model constr. [s]	0.48	0.48	0.47	0.48	0.47	0.47	0.47				
Time of model calc. [s]	10.05	9.60	8.33	9.41	10.31	10.16	9.20				
Nr. of brds. without dep.	4	2	2	2	1	1	0				

Table 1 Number of running boards without depot visit according to iterations of alg.

5 Conclusion

We have defined problem of vehicle scheduling with condition that every running board contains exactly one roundabout route to depot. This problem was formulated as MILP mathematical model followed by a sequence MILP models extended by conditions prohibiting infeasible running boards. We have computed running boards with roundabout routes for public transport system in Slovak town Martin–Vrútky. Computation experiment showed that this model can be applied for real world instances of vehicle scheduling problem to find time positions of roundabout routes to depot.

Compared to our procedure from [5] this model assigns to every running board exactly one dummy trip.

Further research will be focused on ways how to implement also minimization of the cost of running board and maybe other real word constraints on running boards.

Authors of this paper have taken part in bus public transport optimization of more than twenty Czech and Slovak towns, cooperation with everyone of them was only temporary.

Acknowledgements

The authors are thankful to the unknown reviewer for carefully reading this paper and for correcting many unpleasant errors.

The research was supported by the Slovak Research and Development Agency grant APVV-14-0658 "Optimization of urban and regional public personal transport".

References

[1] Ceder, A.: Public-transit vehicle schedules using a minimum crew-cost approach. In: *Total Logistic Management* **vol. 4**, (2011), 21–42.

[2] Gurobi Optimization Inc.: Gurobi Optimizer Reference Manual, *http://www.gurobi.com* (2015).

[3] Palúch, S.: Two approaches to vehicle and crew scheduling in urban and regional bus transport. *Quantitative Methods in Economics (Multiple Criteria Decision Making XIV) Proc. of the Int. Sci. Conf., June 5 – 7, 2008 High Tatras, Slovakia. Bratislava: Iura Edition* (2008), 212–218.

[4] Palúch, S.: Vehicle and crew scheduling problem in regular personal bus transport - the state of art in Slovakia. *16th international conference on transport science – ICTS 2013, 27. Maj 2013, Portorož, Slovenija* (2013), 297–304.

[5] Palúch, S., Majer, T. : Vehicle Scheduling with Roundabout Routes to Depot. *Mathematical methods in economics MME 2016 - 34th international conference, Liberec, Czech republic*, (2016), 623-628.

[6] Peško, Š.: Flexible bus scheduling with overlapping trips. *Quantitative Methods in Economics (Multiple Criteria Decision Making XIV) : Proc. of the Int. Sci. Conf. June 5 - 7, 2008 High Tatras, Slovakia. - Bratislava: Iura Edition* **vol.** (2008), 225–230.

[7] Peško, Š.: The minimum fleet size problem for flexible bus scheduling. *Studies of the Faculty of management science and informatics University of Žilina.* **vol. 9**, 75–81.

Exchange rate and economic growth in Czech Republic

Václava Pánková[1]

Abstract. The aim of the paper is to study a possible impact of the exchange rate on the GDP in the Czech Republic. The relationship between both variables is constructed in the framework of aggregate demand and aggregate supply equations. In 2014 the Czech National Bank started to weaken the exchange rate of the Czech crown with the goal to maintain an economic growth. The effect of such an economic policy is measured by the help of relevant econometric techniques using the annual data. During the years under exchange rate targeting, the GDP growth was higher than before, but without an evidence that the cause is just the exchange rate devaluation.

Keywords: macroeconomic targeting, aggregate demand, aggregate supply, econometric model, stationarity, Marshall – Lerner condition.

JEL Classification: E5, C5
AMS Classification: 62P20

1 Introduction

The aim of the paper is to test the relationship between the exchange rate and GDP in the Czech Republic. In 2014 the CNB (Czech National Bank) started to weaken the exchange rate of the Czech crown because of a disproportionately low inflation rate. The goal of the changing policy was to maintain an economic growth.

Some sort of macroeconomic targeting is applied in many economies, with the goal of generating macroeconomic benefits, in particular to maintain positive economic growth, low inflation, full employment, the satisfactory balance of payments a.s.o. The choice of a target variable is important because there is no direct relationship between monetary instruments and the ultimate goal of monetary policy. A nominal target can affect expectations in two ways. First, it influences market expectations; second, it should encourage firms and workers to behave in a way which seems to be optimal according to their own criteria. The application of a rule is a process the optimality of which, although initially intended, may change under non-optimal conditions. The phenomenon is known as a dynamic inconsistency and is described, for example by Fischer [7]. It also justifies a discretionary policy, explained as making short-term optimal decisions whenever necessary under rational expectations.

In the 1990s, the inflation rule was clearly preferred by central banks; in the CR, inflation targeting was undertaken by the CNB till 2013. Unfortunately, the role of inflation in macroeconomic models is usually simplified; in fact, it is much more complex and stable inflation accompanied by a varying production gap may occur (for a rationale, see Blanchard et al. [4]). After the period of financial crisis, it was clear that stable inflation may be a necessary macroeconomic condition, but it is not sufficient. The 2007–09 financial crisis was clearly a milestone in relation to the macroeconomic targeting concept and gave rise to important conclusions (e.g. [4], [9]), namely that much of the science of monetary policy remains intact. However, the crisis has made it clear that some major rethinking concerning the details of this basic framework for monetary policy strategy is required. So, the nominal GDP targeting, which in fact is 10-year older than the inflation rule concept, was revoked. Also the exchange rate rule became a subject of new interest of central bankers.

The exchange rate rule, which involves holding the exchange rate fixed, is not of great interest to less trade-dependent economies. However, it may bring considerable benefits to a small open economy that is highly dependent on trade. After all, it has successfully been practised in Singapore on a long-term basis [12]. On the other hand, the pitfalls of exchange rate targeting are demonstrated by Belongia and Chrystal, using the case of the United Kingdom (UK) [2].

Recent developments indicate that the long-term application of one rule only will rather be an exception, which has stimulated various authors to look for quantitative comparisons of different policy rules (e.g. [3], [6], [10], [11]). For the study of consequences of alternative policy rules, econometric models are useful.

2 Model and data

The economic activity in CR is represented by GDP; exchange rate CZK/EUR is not only an indicator for the level of country competitive globally but here also a tool of targeting policy. The relationship between both vari-

[1] University of Economics, Faculty of Informatics and Statistics, Dept. of Econometrics, 13067 Praha 3, nám. W. Churchilla 4, pankova@vse.cz

ables is constructed, similar to [1], in the framework of aggregate demand (AD) and aggregate supply (AS) equations.

The demand side

Starting by the identity GDP equals *consumption of the household sector* (C) + *spending on the private sector investment* (I) + *public sector consumption* (G) + *foreign trade balance* (TB), we may reflect that C is a function of income and interest rate, I depends on interest rate and TB is a function of exchange rate ER. Besides, the money supply M is used to replace an interest rate as well as an inflation rate expected. So the AD equation is

$$GDP = \alpha_0 + \alpha_1 M + \alpha_2 ER + \alpha_3 G + v \ . \tag{1}$$

The supply side

The original resources contributing to GDP are labour, capital and other factors of production (dropped in further considerations). Labour is a function (reciprocal) of wages (W), demand for capital depends on an interest rate including an inflation rate expected what is proxied by money supply M again. Dealing with a small open economy, we have in mind that exchange rate plays an important role e.g. in the cost of raw materials. Hence, the AS equation is formulated as

$$GDP = \gamma_0 + \gamma_1 W + \gamma_2 M + \gamma_3 ER + w \ . \tag{2}$$

Paying attention to all important determinants of GDP, the econometric model proposed is

$$GDP = \beta_0 + \beta_1 M + \beta_2 W + \beta_3 ER + \beta_4 D + \beta_5 G + u \tag{3}$$

with D being a dummy variable used to distinguish years 2014 and 2015 during which the exchange rate targeting was applied. This economic policy of CNB was performed also in 2016 and in the beginning of 2017 when a near ending was declared but relevant data were not available in the time of processing this analysis.

The source of annual data used was the Eurostat database. Starting 1995 and ending 2015 inclusive (21 observations), GDP, W, G and M1 are measured in millions of Euro. As ER, the CZK/EUR exchange rate is used.

3 Computations and results

First of all, the question of stationarity of all variables should be solved. The time series in question are relatively short (21 observations), that is why two tests were applied to have a comparison: ADF and KPSS. The results are summarized in Table 1.

variable	ADF-prob	KPSS –LM stat
GDP	0.805	0.605
ΔGDP	0.007	0.148
M	0.981	0.621
ΔM	0.002	0.174
W	0.800	0.603
ΔW	0.005	0.153
ER	0.703	0.546
ΔER	0.000	0.145
G	0.881	0.600
ΔG	0.017	0.168

Table 1 Tests of stationarity

In Table 1, ADF – probability relates to the H_0: time series has a unit root, treated by the t – statistic; KPSS – LM stat is a statistic to be compared with critical values 0.739 – 1 % level, 0.463 – 5 % level, 0.347 – 10 % level under the the H_0: time series is stationary.

According to both tests, all variables are I(1). That is why a cointegration of variables could be tested straightforward using residuals of the regression of (3) and the test of Engle and Granger (EG, see e.g. [5]) based on a

statement that a linear combination of variables is stationary if and only if the $\hat{u}$ is stationary. The regession is described in the Table 2 with

.

Variable	Coefficient	Std. error	t-Statistic	Prob.
M	-0.060909	0.048057	-1.267425	0.2243
W	3.426704	0.263251	13.01688	0.0000
ER	-676.8560	291.8313	-2.319340	0.0349
G	-0.170894	0.180569	-0.946421	0.3589
D	3910.376	1801.182	2.171005	0.0464
C	29809.85	11838.39	2.518067	0.0236

Table 2 Estimation of equation (3)

Applying ADF test to residuals, the statistic is -3.716 what allows to reject H_0: $\hat{u}$ has a unit root at the 5 % significance level (critical value is -3.020) following instructions in [5]. Nevertheless, in [8] the critical values of EG other than those of ADF are recommended including, besides other things, also the number of time series entering the regression. The published critical values relate to large data samples only that is why (i) this recommendation could not be used here, (ii) to support the above conclusions, the Johansen cointegration test was performed. The number of cointegration equations found is non-zero and varies from 2 to 3 according to (non)comprising intercept and / or trend.

Testing for autocorrelation by Breusch – Godfrey test, the $\chi^2(1)$ statistic is 3.783 with Prob = 0.052 what very closely may justify non rejection of H_0: no autocorrelation (in fact, $\hat{\rho} = 0.22$).

Further, the ECM version of (3) was estimated as in Table 3 with D – prefix instead of Δ, R for the error correction part and R-squared =0.986083.

.

Variable	Coefficient	Std. error	t-Statistic	Prob.
DM	-0.035174	0.047038	-0.747781	0.4679
DW	3.290109	0.160415	20.50994	0.0000
DER	-321.5484	185.4593	-1.733795	0.1066
DG	-0.262823	0.124738	-2.106998	0.0551
D	2272.453	851.9870	2.667239	0.0194
R	0.671324	0.213365	3.146366	0.0077
C	463.1251	412.4719	1.122804	0.2818

Table 3 Estimation of the ECM model

The evidence in both Table 2 and Table 3 is that there is no direct influence of the money supply as well as of government spending. On the other hand, the dominating factor influencing GDP evidently are the wages. It is not surprising if we have in mind that the higher wages, the higher consumption is possible, and that the consumption growth is desired to support the GDP growth. The positive development of GDP during the years in which the exchange rate was performed is evident. Nevertheless, it does not explain the role of exchange rate itself. For a nearer explanation let us study the Marshall-Lerner condition, which states that a currency devaluation will only lead to an improvement in the balance of payments if the sum of demand elasticity for imports and exports is greater than one. Provided that the sum of the price elasticity of demand coefficients for exports and imports is greater than one then a fall in the exchange rate will reduce a deficit and a rise will reduce a surplus.

Involving the data of exports (*exp*) and imports (*imp*) (Eurostat; 1995 - 2015 inclusive, annual; in millions of Euro) together with GDP and exchange rate (*gdp er*), all in logarithms we have

$$\hat{exp} = -4.96 + 1.44\,gdp - 0.19\,er$$
$$t = (-1.16)\ (0.17)\quad (0.71)\qquad R^2 = 0.99 \tag{4}$$
$$\hat{imp} = -4.51 + 1.36\,gdp - 0.09\,er$$
$$t = (1.25)\ (0.16)\quad (0.61)\qquad R^2 = 0.98 \tag{5}$$

after treating autocorrelation by the help of GLS. Though an eventual spurious regression might make the values of parameters too high, it is evident that the exchange rate parameters are statistically equal to zero. Hence, the sum of those parameters is less than one. So, we cannot admit any positive influence of the past devaluation.

4 Conclusion and further research

The relationship between the exchange rate and GDP in the Czech Republic was studied respecting the fact of devaluation performed by National bank since 2014. The straightforward calculations show that under the exchange rate targeting the GDP was positively influenced. But an exact designation of the impact variable is out of the ability of the model (3); the influence of exchange rate is not distinct enough, the influence of ages dominates evidently. Employing Marshall-Lerner condition in the form (4) and (5), any positive impact of exchange rate targeting was excluded.

Like similar analyses, this approach also has its weak points. The data concerning 2016 unfortunately was not yet available though, during this year the devaluation was also performed. The Marshall-Lerner condition itself deals with exchange rate but the real effective exchange rate might offer more exact results what refers also to equations (1) – (3). Unfortunately, this data was not disposable as well as alternative (to M1) money aggregates which could allow for variant computations.

Even if accomplishing all the tasks specified above, the apparatus does not allow for distinguishing among different macroeconomic targeting rules. Such a methodology, concerning inflation, respective GDP respective exchange rate targeting is given e. g. in [3] and applied to the Czech economy by Pánková [10], the results indicate that the nominal GDP rule is apparently superior to inflation targeting, which in turn predominates over the exchange rate rule.

Acknowledgements

Financial support of IGA VŠE F4/73/2016 is gratefully acknowledged.

References

[1] Alawin, M., Sawie, K., Al-Omar, I. and Al-Hamadi, M.: Econometric Analysis for the Impact of the Real Effective Exchange Rate on the Jordanian Economy, *European Scientific Journal* **25**, (2013), 69 – 85

[2] Belongia, M. T. and Chrystal, K. A.: *The Pitfalls of Exchange Rate Targeting: A Case Study from the United Kingdom*, 1990. Retrieved from https://research. stlouisfed.org/publications/review/90/09/ Pit falls_Sep_Oct1990.pdf.

[3] Bhandari, P. and Frankel, J.: Nominal GDP Targeting for Developing Countries, *NBER Working Paper* 20898, 2015. Retrieved from http://www.hks.harvard.edu/fs/jfrankel/NGDPT-%20India PBJF2015.pdf

[4] Blanchard, O., Dell´Ariccia, G. and Mauro, P.: *Rethinking Macroeconomic Policy*, International Monetary Fund, 2010. Retrieved from https://www.imf.org/external/pubs/ft/spn/2010/spn1003.pdf.

[5] Charemza, W. and Deadman, D.F.: *New Direction in Econometric Practice*, Edward Elgar Publishing, Massachusetts, 1997.

[6] Edwards, S.: The Relationship between Exchange Rates and Inflation Targeting Revisited, *NBER Working Paper* 12163, 2006. Retrieved from http://www. anderson.ucla.edu/faculty/sebastian.edwards/ fx_inflation.pdf.

[7] Fischer, S.: Rules versus Discretion in Monetary Policy, In: *Handbook of Monetary Economics*, Vol. II, Ed.: Friedman, B. M. and Hahn, F.H., Elsevier, 1990.

[8] MacKinnon, J. G.: Critical Values for Cointegration Tests, *Queen's Economics Department Working Paper* No. 1227 (1990, with new appendix 2010). Retrieved from http://qed.econ.queensu.ca/ working_papers/ papers/qed_wp_ 1227.pdf.

[9] Mishkin, F. S.: Monetary Policy Strategy: Lessons From the Crisis, *NBER Working Paper* 16755, 2011. Retrieved from http://www.nber.org/papers/w16755.

[10] Pánková, V.: GDP Targeting. Econometric Approach. In: *Mathematical Methods in Economics* 15 (MME) Cheb, University of West Bohemia, 2015, 602–605.

[11] Pétursson, T. G. *Exchange Rate or Inflation Targeting in Monetary Policy?*, 2015. Retrieved from http://www. sedlabanki.is/uploads/files/mb001_6.pdf.

[12] Tan S. E.: *Exchange Rate Targeting as a Monetary Policy Alternative to Interest Rate Targeting*, Stanford University, 2008. Retrieved from https://economics. stanford.edu/sites/default/files/publications/ tan_s.e._ 2008_honors_thesis.pdf.

Modelling the Development of the Consumer Price of Sugar

Marie Pechrová[1], Ondřej Šimpach[2]

Abstract. Sector of sugar production and hence production of sugar beet are so far the most regulated branch of the crop production not in the European Union, and hence it is difficult to predict their development. However, knowledge of future consumer sugar prices can help the manufacturers when deciding on the amount of production and during planning. The aim of the article is to predict future development of consumer prices of sugar crystal in the Czech Republic for the period from 09/2016 to 08/2017 using ARIMA model based on the data base of the monthly consumer price of sugar in the period 09/2000–08/2016. Data were obtained from www.agris.com. On the basis of diagnostic tests ARIMA model (1, 1, 1) was selected for modelling. Medium variant forecast predicts moderate growth rates of sugar crystal from 18.40 CZK / kg to 19.12 CZK / kg. The upper variant with a 90% probability promises a significant increase to 22.98 CZK / kg. The low variant expects decline of the price to 15.26 CZK / kg in 08/2017. This is the most probable scenario taking into account the development on the sugar market.

Keywords: ARIMA model, sugar crystal, prediction, consumer price of sugar

JEL Classification: C61, C63
AMS Classification: 62H12

1 Introduction

Sector of sugar production and hence production of sugar beet are so far the most regulated branch of the crop production in the European Union (EU). The quantity produced is limited by quotas, as well as the purchase price is also a subject of adjustments. On the cultivation of sugar beets are provided "subsidies within the scheme of sensitive commodities, which amount to 450 billion CZK (approximately one-third of the original amount of subsidies for sugar beet)" [11]. On imports from third countries are imposed high import duties. Toward the end of sugar quotas and termination of regulation of its production, imports and exports in October 2017, the farmers fear a decline in purchase prices of sugar beet. Sugar factories, on the other hand, welcome that they will be able to decide about the amount of sugar produced and sold by themselves. Currently, the sugar in the Czech Republic (CR) is produced only 7 factories, while 25 years ago there were more than 50 and 150 years ago even 400. Decline of the sugar industry was caused by number of historical events, from the nationalization, the concentration of factories in large units and state intervention in the 70s, through the return to a market economy in the 90s till entrance of the CR to the EU in 2004. Those events lead to "important changes of the owner structure of the land and factories for sugar production, to their closing and consequent changes in distribution channels of the sugar, to massive decline of the sowing areas of sugar beet and strong concentration of planting of the sugar beet in suitable natural conditions, that caused consequent increase of yields of sugar beet" [5]. "Production quota for sugar Czech Republic amounting to 372 459 tons remains in the last fiscal year with quotas at the same level and is divided among five sugar farms with 7 manufacturing plants" [8]. Expected influence of the terminus of quotas on consumer prices of sugar is so-far unclear. The prices of sugar underwent through decline during last 2 years, that caused the decrease of revenues of the sugar factories.

The aim of the paper is to model consumer prices of sugar using econometric methods and predict their development one year ahead. The paper is structured as follows: Firstly, researches that are concerned with price modelling are introduced. Than is describe the Box-Jenkins methodology that is used to model the development of the sugar price. The results are presented in next section and discussed in following one. Last section concludes. Many researches are concerned with modelling of agricultural commodity prices. For modelling volatility of agricultural commodity prices are usually used Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models [17]. To predict the prices is widespread a Box-Jenkins methodology. For example, [10] used ARIMA (1, 0, 0) to model consumer prices of eggs. [12] modelled several agricultural commodities' prices and argued that ARIMA is a suitable method for price modelling. [6] used econometric model to model monthly prices of eggs from March 2000 to September 2009 in China. The explanatory variables were commercial price, feed market prices and first-

[1] Institute of Agricultural Economics and Information, Mánesova 1453/75, 120 00 Prague 2, Czech Republic, pechrova.marie@uzei.cz.
[2] University of Economics Prague, Faculty of Informatics and Statistics, W. Churchill sq.4, 130 67 Prague 3, Czech Republic, ondrej.simpach@vse.cz.

order and second order lag of monthly eggs. For the case of egg prices in Japan, [9] applied autoregressive model and modified multiple regression model on monthly and yearly wholesale egg prices between years 1986 and 1990. Alternative method for time series modelling can be artificial neural network. These models have better statistical characteristics and they can capture a significant number of turning points [4].

2 Methodology

Box-Jenkins methodology was chosen for the purpose of modelling of the time series of the consumer price of sugar. It requires long time series of data. Future development is predicted based on the knowledge of stochastic trend. The model was elaborated in several steps. Firstly, the stationarity of the time series was tested using unit root test. All 3 types of Augmented Dickey-Fuller test (ADF test) – with constant and trend, with constant only, and without constant and trend – were elaborated and was chosen the most appropriate. (In this paper, we have non-seasonal monthly published data, for other cases see e.g. study [14], [15] or [16]). ADF test with constant and trend is calculated as (1).

$$\Delta Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + \sum_{i=1}^{m} \alpha_i Y_{t-i} + \varepsilon_t , \tag{1}$$

where ΔY_t is the first difference of the examined time series, t is trend variable, ε_t is pure white noise error term, m is the maximum length of the lagged dependent variable, and α, β are parameters (β_1 represents the constant). Methodology by [1] was used to elaborate a model with autoregressive (AR) and moving average (MA) processes. Diagnostic of the type of the model is done by Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) that were plotted to determine, what process generated the time series and what is the order p of AR process and order q of MA process. When the time series is not stationary, its difference of d^{th} order must be done. If the time series include all processes it is an ARIMA (p, d, q) model (2).

$$Y_t = \beta + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j} , \tag{2}$$

where α and δ are parameters and β is the constant. When there is a seasonal term, the models are formulated as SARIMA (P, D, Q) (p, d, q). The adjustment was done by algorithm elaborated by Census X13-ARIMA. Verification of the model for autocorrelation was done for by Breusch-Godfrey serial correlation LM test [3]. Null hypothesis states that there is no serial correlation. If the calculated value of the test exceeds the tabled test criterion from Fisher and χ^2 distribution the null hypothesis is rejected and there is autocorrelation. Heteroscedastity was tested by Autoregressive Conditional heteroscedasticity (ARCH) test [3]. Null hypothesis is again that the variance of the stochastic term is constant and finite. The test is also using Fisher and χ^2 critical values and rejects the null hypothesis when the calculated value of the test exceeds the table values. Finally, the normality was tested using Jarque-Bera test [3] with null hypothesis that the residues are normally distributed. Calculated value of the test is compared to critical value of J-B distribution. If the value exceeds the table one, null hypothesis is rejected. Based on the ARIMA model, the predictions are done for next 12 months and 90% confidence intervals are introduced. We used software EViews 8 for the calculations.

3 Data

A total of 192 observations were analysed from 09/2000 up to 08/2016. Data were obtained from the Czech Statistical Office (CZSO) and are available besides others from www.agris.cz. Time series of weekly consumer prices of sugar until 2011 was by calculating the arithmetic mean converted to monthly frequency to match the frequency of data collected from 2011 onwards. Data for 01 and 05/2015 were calculated as the average of previous and following values in the time series.

The average annual sugar prices are given in the Table 1. At the end of 2000 and beginning of 2001, the price was around 22 CZK / kg, but in the second quarter it fell below this level and continued to fall. It fell below 21 CZK / kg in 03/2002 and in the last month of 2002 it fell below 20 CZK / kg. The year 2003 was marked by very low prices, which reached a minimum in 08/2016 (16.76 CZK / kg). Then came slow growth, which peaked price 26.35 CZK / kg the following year. Last quarter of 2004 the price was above 25 CZK / kg. However, in the second quarter of 2005 appeared decline again. It continued in the following year too. But the price did not fall below 21 CZK / kg. The return to growth of crystal sugar consumer price was recorded in 08/2006, when it again exceeded the limit of 22 CZK / kg. In the following months occurred dramatic fluctuations. In 2007 and 2008 the price fluctuated between 20 and 22 CZK / kg. Only in the second quarter of 2009 it dropped below 20 CZK / kg and continued to decline until 01/2011. At the end of 2012 occurred rapid growth on 24.85 CZK / kg. Higher prices

were noted in the following year (on average 24.15 CZK / kg). There was a decline in prices to 19.40 CZK / kg in 12/2014. Prices continued to decline throughout the year 2015 to an absolute minimum in 11/2015. Despite the fact that prices increased this year they reached maximum of only 18.23 CZK / kg in 08/2016.

Year	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Price	22.36*	21.84	20.56	17.74	22.31	23.27	21.73	21.90	21.00
Year	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	
Price	19.81	18.23	22.45	24.27	24.15	22.03	17.75	16.20**	

Note: *Average 09-12/2000, **Average 01-08/2016; Source: own elaboration

Table 1 Average annual consumer prices of sugar crystal (in CZK / kg)

4 Results

Using ADF test without trend and constant it was found out that the time series is non-stationary at 5% significance level (*t*-statistic = -0.6105, p-value = 0.4518). It is stationary after 1st difference, i.e. that the time series is I(1). ACF and PACF analysis revealed that the best model for capturing the time series of sugar is ARIMA (1, 1, 1) without constant. In order to improve statistical characteristics of the model, a unit impulse was added. Variable IMP took value 1 in 08/2004, 02 and 04/2011, where there were shocks in prices, and value 0 otherwise. Using this model, the residues are not autocorrelated (based on the results of *Q*-statistics and Breusch-Godfrey test. The variance of the residues is according to the ARCH test constant and finite. Results of Jarque-Bera test shows that residues are also normally distributed.

Estimated parameters are displayed in Table 2. All parameters are statistically significant (p-value is lower than the level of significance $\alpha = 0.05$). Compared to the alternative models (which results are not displayed here) based on information criteria this model appeared to be the most suitable. Model fulfil all econometric requirements and can be used for predictions. Residual diagnostics are displayed in Table 3 and Figure 1.

Variable	**Coefficient**	**Standard error**	**t-value**	**Probability (p-value)**
IMP	3.2877	0.2878	11.4218	0.0000
AR(1)	0.8347	0.1612	5.1779	0.0000
MA(1)	-0.7386	0.1963	-3.7626	0.0002
R^2	0.4297	Akaike information criterion		1.4399
Adjusted R^2	0.4236	Schwarz criterion		1.4911
Durbin-Watson statistics	2.0662	Hannan-Quinn criterion		1.4606

Table 2 ARIMA(1, 1, 1) model estimates for consumer prices of sugar, Source: own elaboration

Breusch-Godfrey Serial Correlation LM Test:				**Heteroskedasticity Test: ARCH**			
F-statistic	1.1507	Prob. F(2,185)	0.3187	F-statistic	1.9048	Prob. F(1,187)	0.1692
Obs*R-squared	0.7245	Prob. Chi-Sq.(2)	0.6961	Obs*R-squared	1.9058	Prob. Chi-Sq.(1)	0.1674

Table 3 residual diagnostics of ARIMA(1, 1, 1) model, Source: own elaboration

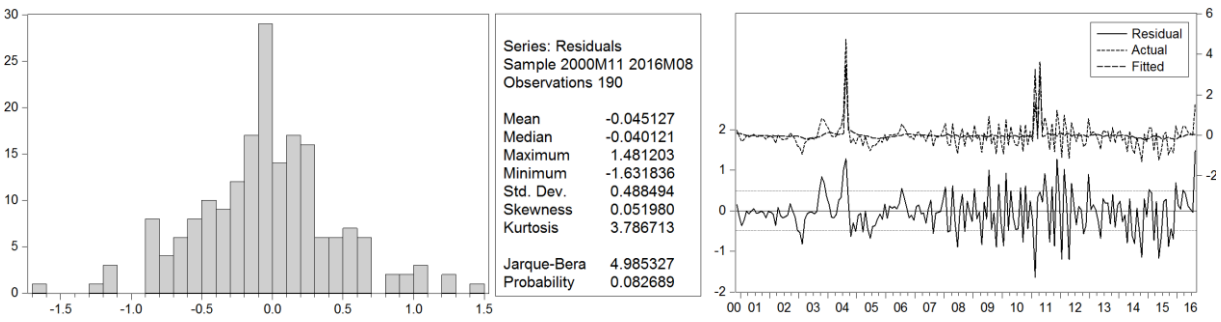


Figure 1 Normal distribution of residuals and graphical control, Source: own elaboration

Development and prediction of consumer prices of sugar, along with a 90% confidence interval is shown in Figure 2. From the graph, there are visible two significant declines in 2003 and again even deeper decline towards the end of 2015. According to the forecast the price should mildly and constantly grow from 17.57 CZK / kg in 09/2016 up to 19.12 CZK / kg at the end of the period 08/2017 without significant fluctuations. This price is relatively low and corresponds to the price, which was on the market between late 2009 and mid-2010. According to the upper variant, there is 90% probability that the growth will be much higher. Price could thus reach 20.64 CZK / kg at the end of 2016 and 22.98 CZK / kg in 08/2017. Still, it will not be reach the high prices that prevailed in the market since the second quarter of 2011 and mid-2014. The lower variant of calculation assumes that with 90% probability the prices will decline from 17.57 CZK / kg in 09/2016 until 15.26 CZK / kg. It reaches nearly a minimum price of 15.17 CZK / kg, that sugar showed at the end of 11/2015. Whereas in recent years the price of sugar has been highly volatile (except for the period from the second quarter of 2011 to mid-2014), for the period 09/2016–08/2017, no significant fluctuations in the price are expected.

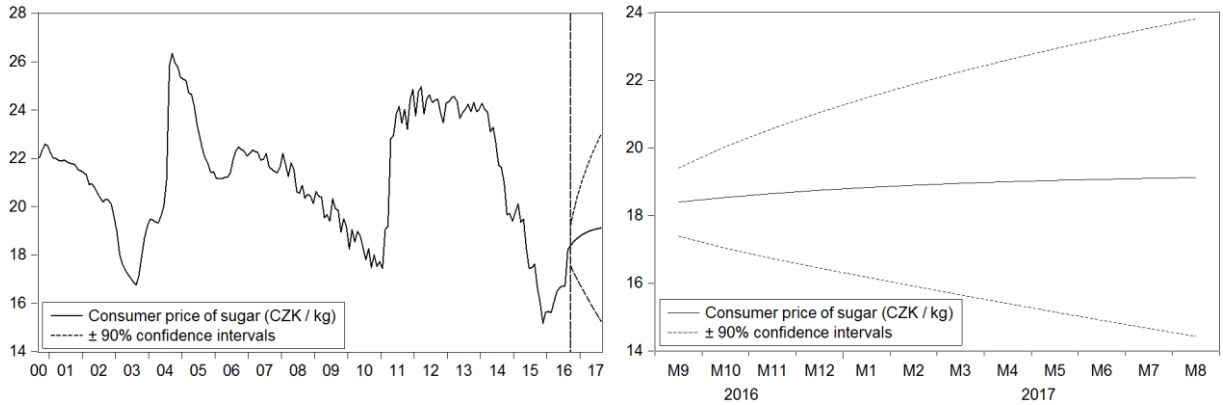


Figure 2 Last development and prediction of consumer price of sugar (CZK / kg), Source: empirical data from www.agris.cz, own elaboration

5 Discussion

Our results can be compared to those of [12]. She also found that the agricultural time series are non-stationary and stationary after first differentiation. For most time series, she analysed, there were also the best type of models ARIMA (1, 1, 1). Similarly, [7] did not find any seasonality in the prices of sugar on the Czech data from year 1993. Regarding the determinants of the price, the analysis of [13] „proved that global sugar price is independent of the level of global sugar reserves level." However, the consumer sugar prices depend on the market situation. As [2] showed the prices of agricultural producers are reflected in consumer prices of sugar.

Ongoing reform of the sugar market in 2006–2009 caused significant decrease of the guarantee price for the producers of sugar beet in EU. This decreased was reflected in real prices, when the prices of sugar producers in 2015 continued in sharp decreased on average 12.13 CZK / kg. In the first half of 2016, the prices were not decreasing any more, but fluctuated around 12.36 CZK / kg of sugar supplied from sugar factories to the market. Consumer prices reaction was fast and the sugar crystal price decreased on average 17.79 CZK / kg during year 2015 and continued to decrease in 2016 (average price for 01–06/2016 was 16.24 CZK / kg). (MoA [8]) Hence, we can expect that the most probable scenario of the price development will be the lower variant – lower bound of 90% confidence interval of our prediction. The price may decrease according to this scenario from 17.57 CZK / kg for 09/2016 to 15.26 CZK / kg in 08/2017.

6 Conclusion

Knowledge of future consumer sugar prices can help the manufacturers when deciding on the amount of production and during planning. Therefore, the article predicted the consumer price of sugar for 09/2016 to 08/2017. Used ARIMA methodology allows to conclude on future price only on the basis of past development of the time series, which is advantageous in terms of data availability. However, the disadvantage is that if it was the past development in time series is very volatile, the prediction may be biased. The aim of the contribution was not only to state what exactly will be the consumer price of sugar in the future, but also to set boundaries in which the price will move with a certain probability. These boundaries are confidence intervals that establish the minimum and maximum price of sugar, which occurs at given time with a 90% probability.

In our case, the predicted price in lower variant was only 17.57 CZK / kg for 09/2016 with a strong trend of decline, falling to 15.26 CZK / kg in 08/2017. This is the most probable scenario taking into account the development on the sugar market. Contrary to that, the upper variant envisages for the same periods growth from 19.22 CZK / kg to 22.98 CZK / kg. Finally, medium variant of forecast expects a slight increase from 18.40 CZK / kg to 19.12 CZK / kg.

Acknowledgements

The research was supported by the Czech Science Foundation project no. P402/12/G097 DYME – "Dynamic Models in Economics" and from Internal Research Grant no. 1110/2017 of Institute of Agricultural Economics and Information.

References

[1] Box, G. E. P., Jenkins, G. M.: *Time series analysis: Forecasting and Control*. San Francisco: Holden-Day, (1970), 537 p.

[2] Hošková, P., Mach, J.: Využití statistických metod při analýze situace na trhu s cukrem [Use of Statistical Methods in Analysis of the Sugar Market Situation]. *Agris.cz – agrární www portál*, (2003), 1–5, URL: http://www.agris.cz/Content/files/main_files/59/137519/hoskova.pdf

[3] Jarque, C. M., Bera, A. K.: Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters*, vol. 6, no. 3, (1980), 255–259.

[4] Kohzadi, N., Boyd, M. S., Kermanshahi, B., and Kaastra, I.: A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, vol. 10, no. 2, (1996), 169–181.

[5] Krejčí, T., Dostál, I., Havlíček, M., and Martinát, S.: Exploring the hidden potential of sugar beet industry brownfields (case study of the Czech Republic). *Transportation Research Part D*, vol. 46, (2016), 284–297.

[6] Li, Z., Li, G., and Wang, Y.: Construction of Short-Term Forecast Model of Eggs Market Price. *Agriculture and Agricultural Science Procedia*, vol. 1, (2010), 396–401.

[7] Mach, J., Hošková, P.: Vliv zahraničního obchodu na domácí ceny zemědělských komodit [The Influence of Foreign Trade Over Domestic Prices of Agricultural Commodities]. *Agris.cz – agrární www portál*, (2005), 192–196, URL: http://www.agris.cz/Content/files/main_files/70/148515/34Mach.pdf

[8] Ministry of Agriculture of the Czech Republic (MoA). *Situační a výhledová zpráva. Cukr – cukrová řepa*. (2006), 1–45, URL: http://www.apic-ak.cz/data_ak/16/k/Cukr/CukrCukrRepaSVZ1606.pdf

[9] Oguri, K., Adachi, H., Yi, C. H., and Sugiyama, M.: Study on egg price forecasting in Japan [1992]. *Research Bulletin of the Faculty of Agriculture - Gifu University*, vol. 57, (1992), 157–164.

[10] Pechrová, M., Šimpach, O.: Modelling the development of eggs prices in the CR. In: *International Scientific Conference on Hradec Economic Days 2017*. Hradec Králové: University of Hradec Králové, (2017), 682–688.

[11] Prośba-Białczyk, U., Sacała, E., and Wilkosz, M.: Vliv listové výživy borem na výnos a technologickou jakost cukrové řepy [Influence of boron foliar fertilization on yield and technological quality of sugar beet roots]. *Listy cukrovarnické a řepařské*, vol. 132, no. 7–8, (2016), 224–227.

[12] Rumánková, L.: Possibilities of Utilization of Univariate Time Series Analysis in Prices Modelling. In: *Agrarian perspectives XXV. – Global and European Challenges for Food Production, Agribusiness and the Rural Economy*. Prague: Czech University of Life Sciences Prague, (2016), 319–326.

[13] Rumánková, L., Smutka, L., Pulkrábek J., and Benešová I.: Vliv zásob cukru na jeho nabídku na světovém trhu [The Influence of Sugar Reserves on Sugar Supply in the World Market]. *Listy cukrovarnické a řepařské*, vol. 128, no. 12, (2012), 381–384.

[14] Šimpach, O., Dotlačilová, P.: The seasonal unit roots in demographic time series and the possibilities of its exact testing. In: *International Scientific Conference on Hradec Economic Days 2013 - Economic Development and Management Region, PT III*. Hradec Králové: Gaudeamus, (2013), 426–432.

[15] Šimpach, O., Dotlačilová, P.: Analysis of Migration Attractiveness of the Regions in the Czech Republic during 2007-2011. In: *Region in the Development of Society 2013*. Brno: Mendel University in Brno, (2013), 322–335.

[16] Šimpach, O., Langhamrová, J.: Development of Socio-Economic Indicators and Mortality Rates during Ten Years of the CR Membership in the EU. In: *Proceedings of the 2nd International Conference on European Integration (ICEI 2014)*. Ostrava: VŠB TU Ostrava, (2014), 675–683.

[17] Yang, J., Haigh, M. S., and Leatham, D. J.: Agricultural liberalization policy and commodity price volatility: a GARCH application. *Applied Economics Letters*, vol. 8, no. 9, (2011), 593–598.

Application of the Measures Based on Quantiles to the Analysis of Income Inequality and Poverty in Poland by Socio-Economic Group

Dorota Pekasiewicz [1], Alina Jędrzejczak[2]

Abstract. Quantiles of income distributions are often applied to the estimation of various inequality and poverty characteristics. They are traditionally estimated using the classical quantile estimator based on a relevant order statistic. The main objective of the paper was to compare the classical and Bernstein estimator for these measures from the point of view of their statistical properties. Several Monte Carlo experiments have been conducted to assess biases and mean squared errors of inequality estimators for different sample sizes under the lognormal or Dagum type-I models. The results of the experiments have been used to the estimation of inequality measures in Poland by socio-economic groups on the basis of the micro data coming from the Household Budget Survey 2014. In the study the head-count ratio, quintile and decile dispersion ratios, maximum equalisation index and the synthetic Zenga index were taken into account.

Keywords: income, inequality, poverty, estimator, bias, mean squared error.

JEL Classification: C13, C15
AMS Classification: 62P25, 65C05

1 Introduction

Statistical measures based on quantiles are frequently applied to the analysis of income distribution as they comprise many popular inequality and poverty indices and indicators. Simple dispersion ratios, defined as the ratios of the income of the richest quantile over that of the poorest quantile, usually utilize deciles and quintiles, but in principle, any quantile of income distribution can be used. A version of the decile dispersion ratio using the ratio of the 10th over the 40th percentile which has recently become popular is the so called Palma Ratio, proposed by Palma in 2011. Another popular inequality measure based on deciles is the coefficient of maximum equalisation, also known as the Schutz index or the Pietra ratio. Contrary to the well-known Gini ratio, the quantile-based dispersion ratios are focused on income differences located in the tails of the distribution rather than in the middle. They can be used as supplementary to overcome the shortcoming of the Gini index of being proportionally oversensitive to changes in the middle of the distribution. More sophisticated measures of income inequality have been constructed using differences (or ratios) between population and income quantiles. Probably the first of such measures was the Holme's coefficient standardized by Bortkiewicz, which is based on the quantiles of order 0.5. The concentration curve and corresponding synthetic concentration coefficient proposed by Zenga, are also defined in terms of quantiles of a size distribution and the corresponding quantiles of the first-moment distribution.

The quantile-based inequality measures are traditionally estimated using the classical quantile estimator based on a relevant order statistic. For some population divisions (by age, occupation, family type or geographical area) these simple methods have been proven seriously biased and the estimation errors were found to be far beyond the values that can be accepted by social policy-makers for making reliable policy decisions (see [4], [6]). The main objective of the paper was to show the advantages of the Bernstein estimator over the classical one for decile and quintile dispersion ratios. After a brief description of quantile-based inequality measures several Monte Carlo experiments have been conducted to assess biases and mean squared errors of their estimators for different sample sizes under the lognormal or Dagum models. Finally, the Bernstein estimator has been applied to the estimation of various income inequality measures in Poland by socio-economic group on the basis of micro data coming from the Household Budget Survey.

[1] University of Łódź/Institute of Statistics and Demography, Department of Statistical Methods, 90-214 Łódź, 41/43 Rewolucji 1905, pekasiewicz@uni.lodz.pl.
[2] University of Łódź/Institute of Statistics and Demography, Department of Statistical Methods, 90-214 Łódź, 41/43 Rewolucji 1905, jedrzej@uni.lodz.pl.

2 Selected statistical inequality measures based on quantiles

One of the most interesting inequality measures that are defined in terms of quantiles is the Zenga index (for details see [10]). The synthetic Zenga index is based on a concentration curve, which, in contrast to the well-known Lorenz curve connected with the Gini index, does not present "forced behaviour" and therefore can be considered a point concentration measure. As a result, both point and synthetic inequality measures proposed by Zenga are sensitive to changes at every "point" of income distribution.

The Zenga point measure of inequality is based on the relation between income and population quantiles ([1], [10]):

$$Z_p = \frac{x_p^* - x_p}{x_p^*} = 1 - \frac{x_p}{x_p^*}, \tag{1}$$

where $x_p = F^{-1}(p)$ denotes the population *p*-quantile and $x_p^* = Q^{-1}(p)$ is the corresponding income quantile. Therefore the Zenga approach consists of comparing the abscissas at which $F(x)$ and $Q(x)$ take the same value *p*.

Zenga synthetic inequality index is defined as simple arithmetic mean of point concentration measures $Z_p, p \in \langle 0, 1 \rangle$. The statistical properties of this coefficient have been discussed in [5].

Other quantile-based inequality measures, much simpler than the Zenga indices, are quintile and decile dispersion ratios (for details see [7]). They can be considered supplementary in income distribution analyses.

The quintile dispersion ratio has the following form:

$$W_{20:20}^{(1)} = \frac{Q_{0.8}}{Q_{0.2}}, \tag{2}$$

where $Q_{0.8}$, $Q_{0.2}$ are quintiles, respectively, the fourth and the first.

The quintile dispersion ratio can also be defined as the ratio of the sum of incomes of the richest 20 percent of the population to the sum of incomes of the poorest 20 percent:

$$W_{20:20}^{(2)} = \frac{\sum_{i \in GK_5} x_i}{\sum_{i \in GK_1} x_i}, \tag{3}$$

where GK_j is *j*-th quintile group.

Using the first and ninth decile we can obtain the following decile dispersion ratio:

$$W_{10:10}^{(1)} = \frac{Q_{0.9}}{Q_{0.1}}, \tag{4}$$

where $Q_{0.9}$, $Q_{0.1}$ are deciles, respectively, the ninth and the first
and

$$W_{10:10}^{(2)} = \frac{\sum_{i \in GD_{10}} x_i}{\sum_{i \in GD_1} x_i}, \tag{5}$$

where GD_j is *j*-th decile group.

The reciprocal of the decile dispersion ratio defined by (5) takes values from the interval [0,1] and is called the dispersion index for the end portions of the distribution:

$$K_{1:10} = \frac{\sum_{i \in GD_1} x_i}{\sum_{i \in GD_{10}} x_i} = \frac{1}{W_{10:10}^{(2)}}. \tag{6}$$

If the index $K_{1:10}$ is closer to the 1, the inequality is lower (mean incomes in the extremal decile groups are the same).

A popular inequality measure based on income shares received by subsequent decile groups is the coefficient of maximum equalisation, also known as the Schutz index or the Pietra ratio:

$$E = \sum_{j \in I} 100 \left(S_j - \frac{1}{10} \right), \tag{7}$$

for $S_j > \frac{1}{10}$ and $S_j = \frac{\sum_{i \in GD_j} x_i}{\sum_{i=1}^{n} x_i}$, where S_j is income share of the *j*-th decile group in the total income. The measure (7) can be interpreted as the portion of the total income that would have to be redistributed (taken from the richer half of the population and given to the poorer half) for there to be income equality.

During a thorough income distribution analysis the problem of inequality measurement is usually interrelated with the estimation of poverty indices. To obtain reliable poverty characteristics it becomes crucial to define and estimate the poverty threshold z_u. There are numerous definitions of this threshold, taking into consideration absolute or relative approach. The relative poverty line utilized by Eurostat is the following:

$$z_u = 0.6M_{0.5}, \tag{8}$$

where $M_{0.5}$ is median of a random variable X.

The poverty threshold can be estimated using the straightforward point estimator: $\hat{z}_u = 0.6Me$, where Me is median estimator established on the basis of a random sample.

Various quantile estimators can be applied to assess the values of Zenga point and synthetic indices as well as those of quintile and decile dispersion ratios or the poverty threshold from sample survey data. In the paper classical quantile estimator and Bernstein estimator are considered.

Classical quantile estimator is defined by the following formula:

$$\hat{Q}_p = \begin{cases} X_{(np)}^{(n)}, & \text{for} \quad np \in N, \\ X_{([np]+1)}^{(n)}, & \text{for} \quad np \notin N, \end{cases} \tag{9}$$

where $X_{(k)}^{(n)}$ is an order statistic of rank *k*.

Bernstein estimator given by:

$$\hat{Q}_p^{Brs} = \sum_{i=1}^{n} \left[\binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \right] X_{(i)}^{(n)}. \tag{10}$$

More examples of quantile estimators, as the standard estimator, Huang-Brill estimator, Harrel-Davis estimator and many others, can be found in [2], [3], [8] and [9]. The analysis presented in the paper are based on the Bernstein estimator, due to its good statistical properties primarily confirmed in the previous study of the authors and presented in [6]. The detailed comparison of the properties of the Bernstein estimator with the classical one are given below.

3 Simulation analysis of quintile and decile dispersion ratio properties

The quantile estimators (9) and (10) were used to the simulation study concerning income inequality measures: $W_{10:10}^{(1)}$ and $W_{20:20}^{(1)}$ given by the formulas (2) and (4). Their statistical properties have been testified under assumed theoretical income distributions presenting different shape and scale parameters. For each combination of the theoretical model, sample size and the type of estimators, a Monte Carlo experiment has been carried out to assess the characteristics of respective sampling distributions.

In the Monto Carlo experiments two different probability distributions were utilized as population models: lognormal distribution, $LG(\mu, \sigma)$, defined by the following density function $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right)$ for $x > 0$ and Dagum distribution $D(p,a,b)$, known also as the Burr type-III distribution, with the density function nn of the form $f(x) = ab^{-ap} p x^{ap-1} \left(1 + \left(\frac{x}{b} \right)^a \right)^{-p-1}$ for $x > 0$. The sets of parameters of both theoretical distributions were established on the basis of real income data coming from Polish HBS and administrative registers, comprising large variety of subpopulations differing in the level of income inequality, which have been observed over the last two decades. The sample sizes were fixed for each variant as n=500; n=1000, n=2000. The number of repetitions of Monte Carlo experiments was N=20 000. The simulated sample spaces were used to assess, for each estimator, its empirical bias and standard error.

The root mean squared errors (RMSE) and relative biases (BIAS) of quintile and decile dispersion ratios have been demonstrated in tables 1 and 2 for different income distributions models and selected sample sizes. All the values are presented as percentages relative to their corresponding population parameters.

Distribution	***n***	**Quintile Dispersion Ratio**			
		$\hat{W}^{(1)}_{20:20}$ **(stand.)**		$\hat{W}^{(1)}_{20:20}$ **(Bernstein)**	
		BIAS	***RMSE***	***BIAS***	***RMSE***
LG(8.0, 0.6)	500	0.113	4.724	0.054	4.532
	1000	0.070	3.335	0.038	3.239
LG(8.1, 0.7)	500	0.063	5.484	-0.014	5.265
	1000	0.011	3.868	-0.017	3.756
LG(8.3, 0.8)	500	0.134	6.252	0.042	6.004
	1000	0.068	4.444	0.027	4.308
D(0.7, 3.6, 3800)	500	0.156	4.454	0.156	4.454
	1000	0.072	3.154	0.072	3.154
D(0.8, 3.0, 3200)	500	0.174	4.978	0.125	4.794
	1000	0.073	3.521	0.053	3.417
D(0.7, 2.8, 3800)	500	0.235	5.753	0.136	5.505
	1000	0.080	4.044	0.046	3.930

Table 1 Properties of Quintile Dispersion Ratio based on quantile estimators

Distribution	***n***	**Decile Dispersion Ratio**			
		$\hat{W}^{(1)}_{10:10}$ **(stand.)**		$\hat{W}^{(1)}_{10:10}$ **(Bernstein)**	
		BIAS	***RMSE***	***BIAS***	***RMSE***
LG(8.0, 0.6)	500	0.126	6.174	0.021	5.882
	1000	0.065	4.327	0.017	4.191
LG(8.1, 0.7)	500	0.124	7.197	0.013	6.868
	1000	0.084	5.088	0.019	4.926
LG(8.3, 0.8)	500	0.186	8.134	0.029	7.758
	1000	0.124	5.815	0.037	5.615
D(0.7, 3.6, 3800)	500	0.353	6.671	0.181	6.344
	1000	0.162	4.702	0.082	4.543
D(0.8, 3.0, 3200)	500	0.347	7.354	0.234	7.002
	1000	0.097	5.179	0.039	5.009
D(0.7, 2.8, 3800)	500	0.554	8.598	0.283	8.181
	1000	0.181	6.003	0.066	5.800

Table 2 Properties of Decile Dispersion Ratio based on quantile estimators

Analysing the results of the calculations it becomes obvious that the Bernstein estimator performs better than the classical estimator of quantile. Its biases and mean squared errors turned out to be substantially smaller for most cases.

4 Application

The inequality measures based on deciles and quintiles, as well as the Zenga indices, have been applied to the inequality analysis in Poland by socio-economic group based on HBS sample 2014. They include the decile and quintile dispersion ratios, the reciprocal of the decile dispersion ratio *K*, the coefficient of maximum equalisation *E* and the synthetic Zenga index *Z*. Basic characteristics of the HBS sample, divided by socio-economic group, are presented in table 3. Table 4 shows the results of the approximation of the empirical distributions by means of the Dagum model. The main outcomes of the application, containing the estimated values of abovementioned inequality measures obtained on the basis of implementation of the Bernstein estimator, are given in table 5.

The relative poverty threshold established as 60% of equivalent national median income and estimated by means of Bernstein estimator is 1181.85 PLN. The estimates of At-Risk-of-Poverty rates (head-count ratios) for each socio-economic group based on this threshold are presented in table 6, In the table there are also the poverty thresholds estimated for each socio-economic group separately.

Socio-economic group	Number of households	Minimum	Maximum	Average	Standard Deviation
Employees	18077	3.00	155017.49	4375.61	2863.20
Farmers	1526	1.67	126739.54	5365.59	6768.69
Self-employed	2473	12.00	42500.00	5358.11	3709.47
Pensioners	13151	13.00	15497.83	2653.50	1487.35
Non-earned sources	1702	20.00	13784.04	1908.38	1401.93
Total	36929	1.67	155017.49	3755.33	2959.95

Table 3 Numerical characteristics of income in socio-economic groups

Socio-economic group	Dagum distribution parameters			Overlap measure
	p	*a*	*b*	
Employees	0.838	3.599	4112.468	0.992
Farmers	0.550	2.407	5298.089	0.976
Self-employed	0.810	3.296	4967.545	0.970
Pensioners	1.283	2.999	2037.633	0.973
Non-earned sources	0.614	3.154	2016.598	0.989
Total	0.747	3.125	3611.017	0.975

Table 4 Approximation of income distributions for socio- economic groups by means of the Dagum model.

Socio-economic group	$W^{(1)}_{20:20}$	$W^{(2)}_{20:20}$	$W^{(1)}_{10:10}$	$W^{(2)}_{10:10}$	$K_{1:10}$	E	Z
Employees	2.311	4.353	3.609	6.689	0.149	20.396	0.255
Farmers	4.094	14.384	10.278	34.918	0.029	34.377	0.586
Self-employed	2.380	5.169	4.061	8.216	0.122	22.594	0.303
Pensioners	2.565	4.416	3.858	6.189	0.162	21.059	0.300
Non-earned sources	2.917	6.812	5.526	11.716	0.085	24.893	0.359
Total	2.819	5.916	4.843	9.526	0.105	22.000	0.345

Table 5 Estimated inequality measures for socio-economic groups

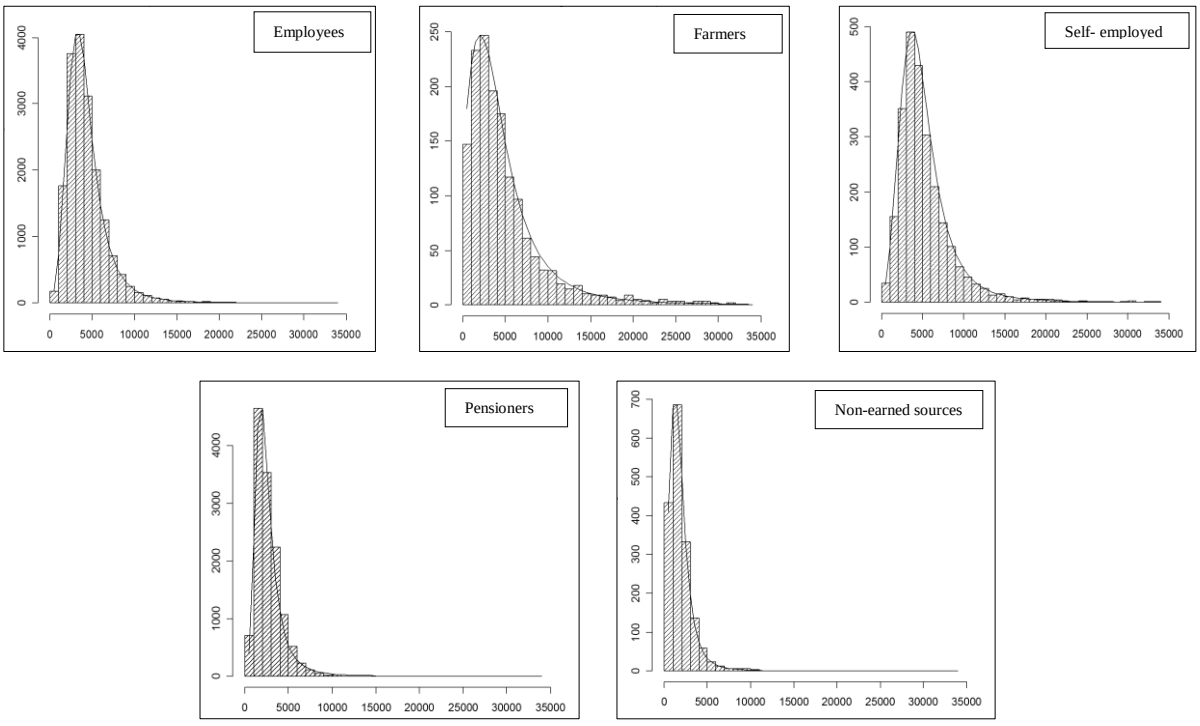


Figure 1 Income distributions for socio-economic groups and fitting by means of the Dagum model

Socio-economic group	Employees	Farmers	Self-employed	Pensioners	Non-earned sources
Head-count ratio	7.39	24.71	5.78	19.63	53.06
Poverty line	1385.30	1208.71	1560.65	1019.35	764.48

Table 6 Estimated poverty lines and head-count ratios [in %] for socio-economic groups

Analysing the results of the approximation presented in table 4 one can easily notice that the Dagum distribution can be applied as a theoretical model describing empirical income distributions in Poland by socio-economic group. We can observe very high consistency of the empirical distributions with the theoretical ones what has also been demonstrated in figure 1. The results of the estimation of inequality measures obtained using Bernstein estimator (table 5) reveal substantial discrepancies in income inequality among socio-economics groups of households. The estimated values of quintile and decile share ratios, as well as the values of synthetic Zenga inequality measures, indicate the *farmers* group as the one with the highest income inequality level. Most of this inequality is located within the extremal decile groups as $W_{10:10}^{(2)}$=34.9 what is very high comparing to the national value equal to 9.5. This group has also high at-risk-of-poverty rate being 24.71% (tab.6). Extremely high value of poverty incidence (53.06%) was observed for the group of households whose primary source of maintenance was *non-earned sources*, comprising mainly unemployment benefits. It is worth noting that high poverty level is not necessarily accompanied by high inequality. In general, 22% of the total income of the Polish households would have to be redistributed from the richer to the poorer groups for there to be income equality (E=22%).

5 Conclusions

The results of the experiments confirm that for different income distributions the Bernstein estimator performs better than the frequently applied classical quantile estimator based on a relevant order statistic. The root mean squared errors for Bernstein estimator are much smaller than those observed for its competitor and its biases are also smaller, especially when the quantiles of higher orders are taken into regard. Consequently, the Bernstein estimator has been applied to the estimation of various inequality measures for socio-economic groups in Poland. Among the quantile-based income distribution characteristics we estimated: decile and quintile dispersion ratios. the reciprocal of the decile dispersion ratio, the coefficient of maximum equalisation, the synthetic Zenga index and the poverty ratio. The analysis revealed substantial discrepancies between socio-economic groups in Poland. The income distribution of *employees* turned out to be most homogeneous while for the households of *farmers* and *non-earned sources* high inequality has been detected, inducing poverty and social exclusion. The obtained results can be the basis of further analysis for economists and social-policy makers.

References

[1] Arcagni, A.: On the decomposition by sources of the Zenga 1984 point and synthetic inequality indexes, *Statistical Methods & Applications*, **26** (2017), 113-133

[2] Harrell, F. E., and Davis, C. E.: A New Distribution-Free Quantile Estimator. *Biometrika* **69** (1982), 635–640.

[3] Huang, M. L., and Brill, P. H.: A Level Crossing Quantile Estimation Method. *Statistics & Probability Letters* **45** (1999), 111–119.

[4] Jędrzejczak, A.: Estimation of Standard Errors of Selected Income Concentration Measures on the Basis of Polish HBS. *International Advances in Economic Research*, **18** (2012), 287-297.

[5] Jędrzejczak, A.: Asymptotic Properties of Some Estimators for Gini and Zenga Inequality Measures: a Simulation Study. *Statistica & Applicazioni*, **13** (2015), 143-162.

[6] Jędrzejczak, A., and Pekasiewicz D.: Analysis of the properties of selected inequality measures based on quantiles with the application to the Polish income data, *Proceedings of the 7th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, in press.

[7] Panek, T: *Ubóstwo, wykluczenie społeczne i nierówności. Teoria i praktyka pomiaru*. Szkoła Główna Handlowa, Warszawa, 2011.

[8] Pekasiewicz, D.: *Statystyki pozycyjne w procedurach estymacji i ich zastosowania w badaniach społeczno-ekonomicznych.* Wydawnictwo Uniwersytetu Łódzkiego, Łódź, 2015.

[9] Zieliński, R.: Small-Sample Quantile Estimators in a Large Nonparametric Model. *Communications in Statistics Theory and Methods* **35** (2006), 1223–1241.

[10] Zenga, M.: Concentration Curves and Concentration Indices Derived from Them. In: *Income and Wealth Distribution, Inequality and Poverty,* Springer -Verlag, Berlin, 1990, 94–110.

A truck loading problem

Jan Pelikán[1]

Abstract. The international transport of goods from manufacturers to customers is in most cases realized by the truck transport. The proposed problem relates to making effective use of these vehicles, considering the way of how to load the trucks effectively and when to send them off. The problem is solved for a particular time interval, during which all requirements for transport are known. Apart from this, the question of whether it is efficient to make use of the cross-dock or the direct routes to customers is discussed. This is dependent on the trade and technical conditions of a particular logistic company. The space for optimization arises thanks to the two options of the transport (direct or via cross-dock) and a possibility that a parcel can be kept not sent off over a certain period of time. The problem is based on a case study. The data from the case study are used for the purposes of computational experiments on the model, verification of the properties of results and their applicability.

Keywords: *truck transportation, cross-dock, integer programming*

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

This article deals with the logistic problem of truck transport, which is commonly used in international transport. Logistic companies provide the truck transport services under particular carriage and trade conditions. A key role in goods transportation plays the cross-dock – a logistic center where the transported goods are transferred from one truck to another, aiming to maximize the usage of the trucks' capacity and minimize the costs of the transport to customers. Using of logistic centers (cross-docks) represents a form of transport, in which parcels are in the first step transported to the logistic center and sorted by the final destination and in the second step loaded onto the trucks heading to these destinations. The trucks leaving from a logistic center then contain parcels from different suppliers, resulting in reduction of costs. A whole truck is used for the transport to cross-dock with fixed costs and not depending on its load volume. On the contrary, the costs for transport from the cross-dock are derived from the weight of goods and the distance covered, as the load consists of goods from diverse suppliers.

The transportation requirements arise at the manufacturer that needs to realize the transport of its goods through a logistic company to the purchasers, trying to minimize the costs during this process. Optimization in a real case study relates to a given time interval during that a list of parcels is set. For every parcel the following values are known: the time when a parcel is ready to be transported, its volume, weight and place of delivery, or if need be other attributes like the priority of the parcel. The transport of the parcel can be carried out within the time period beginning with the day when it was passed on to the logistic company and lasting for the next three days (this is not valid in case of high priority parcels).

The problem that is the subject to this article is based on a case study. Transportation of manufacturer's production abroad is provided by an external logistic company. The problem lies in optimizing the process of loading up the trucks and the schedule of drives. The loaded trucks can travel either directly to the customers or through the cross-dock. In both cases, the trucks are allowed to carry parcels at most to two final destinations. The costs in case of direct delivery do not depend on the weight of goods transported, the costs of trucks departing from cross-dock are calculated according to the weight of goods and the distance covered, where the dependence of the costs on the weight is not linear.

There are two optimization models presented in this article: The first of them optimizes the transport from manufacturer to customers and cross-dock. This is done by minimizing the number of trucks leaving the company in a certain time interval. The second model minimizes the costs of the transport from cross-dock to customers, also in the given time interval.

There are several papers dealing with the task of truck loading in the scientific literature. Their contents are different from the problems solved in this article. Many of these sources solve the problem of cyclic drives of

[1] University of economics Prague, Fakulty of informatics and statistics,W. Churchill sq. 4, 13067 Prague 3
email: pelikan@vse.cz

the trucks loaded with different types of products, e. g. the oil products, so that the time interval between two deliveries is maximal [1] [3]. Another author [2] proposed a way of periodical loading of carriers with more compartments on the load area where the compartments have various capacity and different types of goods are considered. Yuceer [4] proved this problem to be NP-hard, proposed a mathematical model, analyzed its properties and proposed a solving algorithm.

2. THE OPTIMIZATION MODEL FOR PARCELS TRANSPORTED BY THE TRUCKS FROM THE FACTORY

The time interval <0, *T*> is given along with the number of parcels *n* that should be transported from manufacturer to customers.
For each parcel the time, when it is ready to be transported, is known. If it is a high priority parcel, the time must be respected; in case of ordinary parcels it is possible to postpone their transport over a specific number of days (in our model we assume the delay of two days).
We know the volume and the weight of a parcel. The company providing the transport offers two options: The direct delivery to a customer, or via the cross-dock. The transport is realized by trucks of a given capacity. If it is a direct delivery truck (FTL direct), then it can contain parcels assigned to a maximum of two customers.
The aim is to determine the way of transport with the minimal number of trucks, which means to set for each truck its time of getting ready and departing, what will be loaded onto it and whether it is a direct or cross-dock type of truck. The number of available trucks is not limited but for the purposes of formulation of the model we will assume a certain maximal number of trucks every day, this amount matches the need in the case study. The capacity of each truck is 26 pallets, the load bearing capacity is not a constraining factor.

The mathematical model contains the following parameters and variables.

Parameters of the model:
t_i the time when the parcel *i* is ready for transport,
q_i the volume of the parcel *i*, measured in the number of pallets,
w_i the weight of the parcel *i*,
$d_i^{\,s}$ the binary parameter, equals to 1 if the *i*-th parcel should be delivered for the destination *s*.

Variables of the model:
x_{ijk} a binary variable, equals to 1 if the parcel *i* is sent off in the time *j* in a direct truck *k*, equals to 0 otherwise,
x'_{ijl} a binary variable, equals to 1 if the parcel *i* is sent off in the time *j* to the cross-dock, equals to 0 otherwise,
y_{jk} a binary variable, equals to 1 if *k*-th truck is sent in the time *j* directly to the customers,
y'_{jl} a binary variable, equals to 1 if *l*-th truck is sent in the time *j* to the cross-dock,
h_{jks} a binary variable, equals to 1 if *k*-th truck goes to *s*-th destination in the time *j*.

The mathematical model:

$$\sum_{j,k} y_{jk} + \sum_{j,l} y'_{jl} \;\rightarrow\; \min \tag{1}$$

$$\sum_i x_{ijk} q_{i\leq} \leq 26 y_{jk}, \quad \forall j,k \tag{2}$$

$$\sum_i x'_{ijl} q_{i\leq} \leq 26 y'_{jl}, \quad \forall j,l \tag{3}$$

$$\sum_{jk} x_{ijk} + \sum_{jl} x'_{ijl} = 1, \quad \forall i \tag{4}$$

$$x_{ijk} = x'_{ijl} = 0, \quad \forall i,k,l \quad \forall j > t_i + 2, \;\; j < t_i \tag{5}$$

$$\sum_s h_{jks} \leq 2, \quad \forall j,k \tag{6}$$

$$x_{ijk} \leq \sum_s d_{is} h_{jks}, \quad \forall i,j,k \tag{7}$$

The object function (1) is a number of all trucks used in transport of the parcels in the given time interval. Inequality (2) assures that if the truck k does not go to the cross-dock in the time j, then it is not possible to use it for transport of the parcels. At the same, this inequality limits the amount of the load by the capacity of the truck. Similarly, (3) is valid for the trucks going directly to customers. Thanks to (5), the parcel is not allowed to be loaded earlier than it is available, or later than it is feasible, which is more than two days later. The inequality (6) expresses the

condition of at most two destinations for the direct trucks. Constraint (7) does not allow loading a parcel into a truck that will not arrive at the destination of the parcel.
The model is linear with binary variables, but it can contain a large number of binary variables in the real world instances in case of a big number of parcels and a long time horizon, even though the condition (5) fixes many of them to zero.
The object function (1) is the sum of all tracks outgoing the factory. Another object function can be costs, which are fixed costs and variable costs. Fixed costs can be in the form of sum $\sum_{jk} y_{jk} + \sum_{jl} y'_{jl} + \sum_{jr} y''_{jr}$, where c_1 are fixed costs of one truck going direct from factory to destination, c_2 are fixed costs of truck from factory to cross-dock, and c_3 are fixed costs of truck from cross-dock to destination and y_{jr} is binary variable with value one if truck r goes from cross-dock to destinations in the time j (a part, modelling the transport from cross-dock to destinations, has to add to the model). If only fixed costs are optimized, it is supposed, that the variable costs are constant, independent on vehicle's journey.
To optimize all costs, fixed and variable together, especially if variable costs is non-linear dependent on weight of parcels and length of truck's route, we have to proceed as described in the next chapter.

3. NONLINEAR COSTS OPTIMIZATION MODEL OF TRANSPORT GOODS FROM CROSS-DOCK TO DESTINATIONS

The trucks departing from the cross-dock, drive immediately to customers. Even now it is valid that they can drive to two destinations. Contrary to the first proposed model, we will minimize the costs that are dependent on the weight of parcels, the relation between them being non-linear. The price of one transported unit is given in the table 1.

Interval	1	2		L
Weight[t]	(r_1, r_2)	$< r_2, r_3)$	...	$< r_L, \infty)$
Price 1 tkm	c_1	c_2	...	c_L

Table 1 Costs of one ton-kilometer transported by truck

The parameters from the previous model remain unchanged (except t_i) and this time the following variables and parameter are newly defined:
t_i the time when the parcel i is ready for transport from cross-dock,
x_{ijk} a binary variable, equals to 1 if the parcel i is sent in the time j directly from cross-dock to the customer in the truck k, 0 otherwise,
y_{jk} a binary variable, equals to 1 if k-th truck departs in the time j from the cross-dock to customers,
h_{jks} a binary variable, equals to 1 if k-th truck goes to the s-th destination in the time j.

The constraints (2), (6) and (7) are the same as in the previous model. The constraints (4) and (5) are altered in this way:

$$\sum_{jk} x_{ijk} = 1, \quad \forall i \qquad (4')$$

$$x_{ijk} = 0, \quad \forall i,k,l \quad \forall j > t_i + 2, \quad j < t_i \; \forall i,j,k \qquad (5').$$

It is assumed that the cost for using a particular truck are now dependent on the product of the weight and the distance covered by this truck (tkm), where the price for one tkm is given by the table 1. To incorporate this price into our model, we need to introduce other variables:

$$u_{jk} = \sum_i \; x_{ijk} \, w_i \qquad (8)$$

the weight of k-th truck dispatched from the cross-dock in the time j, a binary variable f_{jkl} that equals to one if the weight of the truck k is in the price interval 1 in the time j. This variable then satisfies the inequality (9) where M is a big number.

$$-M\left(1 - f_{jkl}\right) + f_{jkl} \, r_l \leq u_{jk} \leq \; r_{l+1} f_{jkl} - \frac{1}{M} + M\left(1 - f_{jkl}\right) \; \forall j,k,l \; . \qquad (9)$$

The second part of the costs of the truck *k* transported in the time *j* is the length of its route, which will be denoted as D_{kj}. To calculate it, we will need to know the distance of each destination from the cross-dock, denoted as d_{1s}, and the distance between two destinations *s1* and *s2*, denoted as $km_{s1,s2}$. Then the distance covered by the *k*-th truck in the time *j* can be expressed as (9).

$$D_{kj} = \sum_s km_{1,s}\, h_{jks} + \sum_{s1,s2} km_{s1,s2} h_{jks1} h_{jks2} \,. \qquad (10)$$

Then the goal function summarizing the costs of transport from the cross-dock to customers forms a non-linear function (11).

$$\sum_{k,j,l} u_{jk} c_l f_{jkl} D_{kj} \;\rightarrow\; min. \qquad (11)$$

The mathematical model optimizing the transport from cross-dock to customers is made by

a) a non-linear goal function (11),
b) constraints: (2), (6), (7), (4'), (5'), (8), (9) and (10).

5. NUMERICAL EXPERIMENTS

The described model of operating the international truck transport is based on a case study and the proposed models were verified on the data from a company involved. The time interval for this computation was 14 days, during which approximately 1000 parcels were to be transported to 10 destinations. Mathematical model (1)-(7) contains 60 thousands binary variables and app. 100 thousands constraints. The model was proved on PC (InteCore2Quad, 2,83GHz, CPLEX 12.0). The computation was interrupted after 30 mins. a good feasible solution was obtained with gap 3,5%.

6. CONCLUSION

This article proposes models for optimization of the truck transport. These models are based on a set of parcels given in advance, which should be transported. If the parcels are not known in advance then the optimization is based on a decision whether a parcel ready for the delivery should be loaded onto truck and sent off or, if the truck is not fully loaded, postpone the delivery to the time when the truck is loaded more efficiently with new incoming parcels. The essential matter is then what percentage of the full capacity of the truck is economical sufficient for sending the truck off, remembering that it is not possible to wait too long for the parcels until their sending off. The presented mathematical model and historical data can be used to establish the load limit. As it was stated earlier, the model contains a significant number of variables depending mostly on the number of parcels in the given time interval. Therefore, the applicability of the model will be higher with a smaller number of parcels.

Acknowledgements

Supported by the grant No. 16-00408S of the Czech Grant Agency.

References

[1] Bukchin Y., and Sarin S.C.: A cyclic policy for the loading of multiple products on a vehicle with different compartment sizes. In: *IIE Transactions* **30/7** (2004) pp.641-653.
[2] Brown G.G. and Graves G.W.: Real-time dispatching of petroleum tank trucks. In: *Management Science* **27** (1981) pp. 19-32.
[3] Ronen D.: Dispatching petroleum products. In: *Operations Research* **43** (1994) pp. 379-387.
[4] Yuceer, U.: A multi-product loading problem: A model a solution method. In: *European Journal of Operational Research*, **101**(1997), pp. 519-631.

Competitiveness Evaluation of Czech Republic Regions with Data Envelopment Analysis

Natalie Pelloneová[1], Eva Štichhauerová[2]

Abstract. The article deals with the competitiveness evaluation of Czech Republic regions using the data envelopment analysis. The aim of the paper is to analyze a competitive potential of Czech Republic regions at NUTS 3 level by quantitative efficiency analysis. The object of the research was 14 administrative regions including Prague. The article is divided into three parts. The first part focuses on a literature review of the use of DEA in evaluating regional efficiency. Nowadays, DEA becomes a suitable analysis tool for setting a competitive resp. non-competitive position of each NUTS 3 region within the Czech Republic. The next part of the article describes the research methodology. The method of DEA used for the evaluation is applied to the NUTS 3 regions based on a particular set of regional input and output variables and other regional characteristics. The regional inputs and outputs variables form the key elements of the system evaluated for each administrative region in the sense of its effective resp. ineffective economic status. Regional efficiency can be thus perceived as a reflection of competitiveness. The last part of the article presents the results of the analysis.

Keywords: NUTS 3, competitiveness, efficiency, data envelopment analysis, decision making units, CCR model, BCC model.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The contribution deals with the examination into the competitiveness of regions in the Czech Republic at the NUTS 3 level. The contribution is targeted at the application of a selected method of Data Envelopment Analysis (DEA) to evaluate the efficiency of regions. Using the DEA method, it is possible to evaluate the efficiency with which regions can transform their inputs into outputs. This fact allows us to consider the efficiency of regions to be a certain reflection of competitiveness. When applying the DEA method, our hypothesis is that the higher the level of development and performance of regions is, the higher their rate of efficiency is, and therefore also the competitive potential of the regions is higher [14].

The term competitiveness is used to express the economic advantage of the national economy, a region, or possibly an enterprise within economic competition, and it is currently monitored very carefully [4]. Competitiveness may be defined as the ability of enterprises, industries, regions, nations and multinational regions to generate a high level of income and rate of employment. It is important to differentiate between the competitiveness at the macroeconomic level (the performance of a given economy in relation to another economy) and at the microeconomic level as the competitiveness of an enterprise within a given market. Regions can be understood to be component units of states, which means the mezzo-level [5]. The competitiveness of regions is connected with their ability to generate a high level of value from their own and acquired sources [6]. Therefore, competitiveness has become the main momentum of the development strategy of states and regions [4]. The enhancement of the competitiveness of regions results in increasing their population's standard of living, and enhancing their political-economic positions [17]. Regarding the evaluation of the competitiveness of regions, the key precondition lies in the economic performance of regions, which may be evaluated using GDP per capita, the productivity of labour, and the unemployment rate [10]. For regions at the NUTS 2 level, the regional competitiveness index is used in the European Union.

An important precondition for enhancing production unit behaviour in a competitive environment lies in the measurement of the performance and efficiency of the units. The competitiveness of a region can be defined as its ability to exploit certain inputs to optimally produce outputs. This contribution applies the Data Envelopment Analysis, which is among the group of non-parametric methods used for measuring technical efficiency [11]. DEA allows for making a comparison of the efficiency of the transformation of multiple inputs to multiple outputs [7]. The method is suitable for evaluating the relative efficiency of Decision Making Units (DMUs) that

[1] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, natalie.pelloneova@tul.cz
[2] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, eva.stichhauerova@tul.cz

use the same input items to produce the same output items. It is possible to evaluate various institutions, territorial units, enterprises, etc. [13]. Unit performance is evaluated by comparing its performance with the performance of the best functioning units in the model. Units with the best performance create the efficiency frontier. Regarding inefficient production units, it is possible to discover how a unit should reduce its inputs or increase its outputs in order to become more efficient [2].

DEA models may be generally classified from two viewpoints: their orientation on inputs/outputs, and their base in variable/constant returns to scale [1]. In input-/output-orientated models, attention is paid either to the augmentation of outputs or the contraction of inputs [14]. An input-orientated model minimizes its inputs satisfying the present output level. An output-orientated model maximises outputs maintaining the present input level [9]. A model based on a precondition of constant return to scale is called the CCR model. By combining with the viewpoint of input/output orientation, it is possible to differentiate between an input-orientated CCR model (CCR-I) and an output-orientated CCR model (CCR-O). The CCR-I model helps to determine the amount of input items needed for a unit in order to make an inefficient unit efficient. The CCR-O model, on the contrary, helps to determine the amount of outputs needed in order to make an inefficient unit efficient. The technical efficiency ratio is defined as the ratio between the weighted sum of inputs and the weighted sum of outputs. The weight must be determined in such a way that makes the value of the ratio greater than or equal to 1. If the value of the ratio is greater than one, this means that the unit is not working efficiently, and expresses the need for a proportional increase in the enterprise's output items [12]. A model based on variable returns to scale is called the BCC model, which also exists in two variants: an input-orientated model (BCC-I) and an output-orientated model (BCC-O). The model is unique for its convex data envelope, which results in the fact that efficient units comprise more production units than the CCR model [16]. For the possible further evaluation of all efficient DMUs, the super-efficiency model has been developed. Super-efficiency models measure the distance between the evaluated unit's inputs and outputs from the new efficiency border. In this paper, Andersen and Petersen model (AP model) was applied which was first published in 1993. In super-efficiency models scores of inefficient units remain unchanged but the efficiency score of efficient units may be higher/lower than 1. Efficient units subsequently have, in the case of output-orientated models, ratio values less than one [12].

The advantages of the data envelopment analysis include the possibility of applying it to multiple inputs/outputs and obtaining a single overall score of the relative efficiency for each DMU [2]. The disadvantage lies in the fact that the small number of observations compared with the large number of used criteria increases the number of DMUs at the efficiency border. Therefore, it is recommended that one uses a number of criteria that is at least three times lower than the number of examined DMUs. Selected criteria should be cardinal for DMU performance, and their values must be known for all DMUs. Another disadvantage lies in the fact that the DEA does not address the possible occurrence of extreme values in the criteria. If an extreme point appears in the analysis, it may result in a deviation of the efficiency estimation of the analysed DMU, possibly shift the efficiency border, and deflect the efficiency estimations of all analysed DMUs [11].

2 Data and methodology

The DEA method is applied in the contribution to 14 regions of the NUTS 3 level in the Czech Republic. In view of the small number of production units, it uses models with two or three input and output items. The criteria were selected in such a way that the correlation between them was not too high (i.e. the maximum correlation coefficient value is 0.8). The reference period is determined by the availability of the selected indicators at the regional level from 2011 to 2015 [3].

Two input items were selected for the application of the chosen DEA method (see Table 3). The first input item contains R&D expenses provided. R&D expenses are one of the main components of economic growth, and they support future competitiveness. The other input item contains the number of inhabitants in a given region.

Three output items were used (see Table 3): the first output item contains net household disposable income (NHDI). Net household disposable income describes the standard of living and purchasing power of the population in a given region. The second output item contains GDP per employed person; by means of this indicator, the productivity of labour is expressed, or it expresses what production has been created by economically active populations in a given region, and what the labour efficiency is there. An increase in productivity of labour is generally considered to reflect an enhanced economic level and competitiveness in the region's economy. The third output item contains employees' average gross monthly salary.

The choice of a specific DEA model depends on which of the above-mentioned characteristics can be influenced, and which of them cannot. As it is possible to influence output item values to a large extent, the following section will only deal with output-orientated efficiency models. This concept means that a region with an efficiency ratio equal to 1 is efficient, whereas an efficiency ratio greater than 1 indicates an inefficient region, and determines the rate of increase in outputs necessary to ensure the efficiency of the region.

The CCR-O output-orientated model (with multiple inputs and outputs) was applied on the data. The model was developed with the assumption of variable returns to scale. As a model with constant returns to scale results in a rather small number of efficient DMUs, it is possible to arrange regions fairly precisely. Dual form of CCR-O model can be formulated as:

$$\frac{1}{\mathrm{E}_0} = \max. \varphi + \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right)$$

$$\text{s.t.} \quad \sum_{j=1}^{n} \lambda_j X_{ij} + s_i^- = X_{i0}, i = 1, \dots, m$$

$$\sum_{j=1}^{n} \lambda_j Y_{rj} - s_i^+ = \varphi Y_{r0}, r = 1, \dots, s$$

$$\lambda_j, s_i^-, s_i^+ \geq 0, j = 1, \dots, n, i = 1, \dots, m, r = 1, \dots, s. \varphi \; unrestricted \; in \; sign.$$

where λ_j, $j = 1, 2, \dots, n$ are weights of all DMUs, s^-_i, $i = 1, 2, \dots, m$ and s^+_r, $r = 1, 2, \dots, s$ are slack/surplus variables, φ is the efficiency score that expresses the improvement rate of outputs in order this unit reaches the efficient frontier.

For the purpose of the further classification of efficient regions, the above-mentioned efficiency model was supplemented with Andersen and Petersen super-efficiency model. Therefore, it was possible to create a ranking of all regions from efficient to inefficient based on the results. Its formulation is very close to the standard formulation of the CCR-O model. Only difference is that the weights of the DMU_q, i.e. λ_q, are set to zero: $\lambda_q = 0$. This causes that the DMU_q is removed from the set of units and the efficient frontier changes its shape after this removal [8].

3 Research results

This chapter presents the results of the data analysis. It determines which of the regions are inefficient, and how the individual output items should be modified. The CCR-O model is applied together with subsequent Andersen and Petersen super-efficiency model. The MS Excel accessory DEA-solver was used to solve the matter.

The following Table 1 shows the relative efficiency scores in given regions per individual year, as well as an overall efficiency assessment. Efficient regions are marked in grey in Table 1. The greatest number of efficient units was contained in the CCR-O model in 2014 (five). The lowest efficiency (1.217) was seen in the South Moravian Region in 2012. On the contrary, the Capital City of Prague and the Central Bohemian and Karlovy Vary Regions were always marked as efficient during the five examined periods. Regarding the ranking, the leading positions were held by the Capital City of Prague, and then the Central Bohemian and Karlovy Vary Regions. The South Moravian Region was ranked last.

Region	2011	2012	2013	2014	2015	Aver. efficiency	Ranking
Capital City of Prague	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1.
Central Bohemian region	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1.
South Bohemian region	1,0652	1,0713	1,0609	1,0365	1,0658	1,0599	5.
Pilsen region	1,0452	1,0390	1,0275	1,0415	1,0168	1,0340	4.
Karlovy Vary region	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1.
Ústí region	1,0406	1,0577	1,0607	1,0747	1,0709	1,0609	6.
Liberec region	1,0773	1,0872	1,0807	1,0739	1,0687	1,0776	9.
Hradec Králové region	1,0335	1,0141	1,0127	1,0000	1,0000	1,0120	3.
Pardubice region	1,0973	1,0763	1,0734	1,0479	1,0542	1,0698	8.
Vysočina region	1,0000	1,0000	1,0000	1,0000	1,0106	1,0021	2.
South Moravian region	1,2012	1,2172	1,1852	1,1662	1,1423	1,1824	12.
Olomouc region	1,1164	1,1478	1,1290	1,1165	1,1059	1,1231	11.
Zlín region	1,0862	1,0764	1,0690	1,0609	1,0462	1,0677	7.
Moravian-Silesian region	1,1562	1,1433	1,1027	1,0920	1,1016	1,1192	10.

Table 1 CCR-O model results

The original CCR-O model resulted in the identification of up to five efficient regions in a single period. Based on the average relative efficiency score, there were three efficient regions; the super-efficiency model was used to rank them further**.** Those regions were examined as shown in the following Table 2. In the grey-marked area of Table 2, the efficiency ratio values are different from those shown in Table 1.

Region	2011	2012	2013	2014	2015	Aver. efficiency	Ranking
Capital City of Prague	0,7811	0,7981	0,7927	0,7978	0,8050	0,7950	2.
Central Bohemian region	0,9502	0,9305	0,9872	0,9712	0,9839	0,9646	3.
South Bohemian region	1,0652	1,0713	1,0609	1,0365	1,0658	1,0599	7.
Pilsen region	1,0452	1,0390	1,0275	1,0415	1,0168	1,0340	6.
Karlovy Vary region	0,1672	0,2377	0,1135	0,1313	0,1941	0,1688	1.
Ústí region	1,0406	1,0577	1,0607	1,0747	1,0709	1,0609	8.
Liberec region	1,0773	1,0872	1,0807	1,0739	1,0687	1,0776	11.
Hradec Králové region	1,0335	1,0141	1,0127	0,9977	0,9847	1,0085	5.
Pardubice region	1,0973	1,0763	1,0734	1,0479	1,0542	1,0698	10.
Vysočina region	0,9849	0,9805	0,9804	0,9986	1,0106	0,9910	4.
South Moravian region	1,2012	1,2172	1,1852	1,1662	1,1423	1,1824	14.
Olomouc region	1,1164	1,1478	1,1290	1,1165	1,1059	1,1231	13.
Zlín region	1,0862	1,0764	1,0690	1,0609	1,0462	1,0677	9.
Moravian-Silesian region	1,1562	1,1433	1,1027	1,0920	1,1016	1,1192	12.

Table 2 CCR-O super-efficiency model results

As a result of the super-efficiency ratio determination for the originally efficient regions, a new ranking of all regions was obtained (see the last column in Table 2). The efficient regions were ranked as follows; the analysis showed that the Karlovy Vary Region was the most efficient region, followed by the Capital City of Prague and the Central Bohemian Region. The Olomouc and South Moravian Regions were marked as the least efficient regions.

4 Conclusion

The contribution dealt with the competitiveness of 14 regions of the NUTS 3 level in the Czech Republic. Competitiveness was described as the ability of a region to efficiently transform multiple inputs into multiple outputs. DEA was used, and the relative efficiency of the regions was applied to evaluate the rate of competitiveness. The method was applied to data publicly available in the CZSO database; it contained values of five criteria for 14 regions in 2011–2015. Two input items and three output items were used, and they met the precondition that the correlation coefficient must be lower than 0.8. The output-orientated model was applied: the CCR-O model assuming constant returns to scale, and it was supplemented with Andersen and Petersen super-efficiency model determining the ranking of the efficient units from the original efficiency border. The result of applying models to the data created an overview of the relative efficiency score development in each region for 2011–2015, and the average relative efficiency score for the five-year period based on which a ranking of the regions was created (see Table 2).

Based on the average score, three efficient regions were identified. The Karlovy Vary Region, followed by the Capital City of Prague, were in the first two positions. The Karlovy Vary Region is considered to be a region with a not very favourable competitive position (unlike the capital Prague with its excellent competitive position). Although the absolute input and output values are not high, the ability of efficient transformation, or optimum exploitation of inputs to create outputs, is at a high level in the Karlovy Vary Region.

On the opposite side, the Olomouc and South Moravian Regions were ranked the last two by this model. And the fact is that the South Moravian Region with its regional capital of Brno is often considered a region with a favourable competitive position. The South Moravian Region can be used as an example for a subsequent evaluation of how a region should increase its output to enhance its efficiency. In order to regard the South Moravian Region in 2015 as efficient, taking the results of the CCR-O model's application into consideration, it would be necessary to change the output as follows: to increase the productivity of labour by approximately 33% (i.e. from CZK 876 to 1 168 per employee), the average gross monthly salary by approximately 14% (i.e. from CZK 27 051 to 30 900), and net household disposable income (NHDI) by approximately 14% (i.e. from CZK 249.89 to approximately 285.44 million). If the competitiveness of a region is defined unconventionally as its

ability to optimally exploit certain input items to produce output items, the DEA method is a suitable instrument to create a ranking of regions in the Czech Republic based on the degree of this ability. In the future, the analysis should be followed by an analysis in the economic policy area and an examination into the ways through which it is possible to feasibly influence the values of output items in order to make inefficient regions more efficient.

Acknowledgements

Supported by the TUL - Technical University of Liberec, Faculty of Economics under SGS grant No. 21189.

References

[1] Charles, V., and Kumar, M.: *Business Performance Measurement and Management*. Cambridge Scholars Publishing, Newcastle upon Tyne, 2014.
[2] Charnes, A., Cooper, W. W., Lewin, A. Y., and Seiford, L. M.: *Data Envelopment Analysis: Theory, Methodology, and Application*. Springer Science+Business Media, New York, 1994.
[3] CZSO: Data from Internal Database of Czech Statistical Office. Praha: Czech Statistical Office, 2016.
[4] Černý, J.: The Influence of Triad Participants on the Competitiveness of regions: the Zlín Region in Europe. In: *Proceeding of the 6th European Conference on Innovation and Entrepreneurship* (Fulford, H., ed.). Aberdeen Business Scholl: Robert Gordon University, Scotland, 2011, 190–196.
[5] Dzemydaitė, G., and Galinienė, B.: Evaluation of Regional Efficiency Disparities by Efficient Frontier Analysis. *Ekonomika* **92**, 4 (2013), 21–36.
[6] Elleithy, K.: Advanced techniques in computing sciences and software engineering. Springer, New York, 2010.
[7] Grmanová, E.: Application of Two-Stage DEA Technique for Efficiencies Measuring of Insurance Companies and Banks. *Trends Economics and Management* **7**, 16 (2013), 9–15.
[8] Jablonský, J.: Using analytic hierarchy process as a tool for ranking of efficient units in DEA models. In: *Proceedings of the International Symposium on the Analytic Hierarchy Process 2011* (Petrillo, A., ed.). Creative Decision Foundation on behalf of the International Symposium on the Analytic Hierarchy Process, Pittsburgh, 2011, 1–6.
[9] Ji, Y., and Lee, C.: Data Envelopment analysis. *The Stata Journal* **10**, 2 (2010), 267–280.
[10] Kislingerová, E.: Inovace nástrojů ekonomiky a managementu organizací. C.H. Beck, Praha, 2008.
[11] Kočišová, K.: Využívanie platobných kariet a efektívnosť bánk. *E+M Ekonomie a Management* **14,** 1 (2014), 121–139.
[12] Kresl, P. K., and Sobrino, J.: *Handbook of Research Methods and Applications in Urban Economies*. Edward Elgar Publishing Limited, Cheltenham, 2013.
[13] Mantri, J. K.: *Research methodology on data envelopment analysis (DEA)*. Universal-Publishers, Boca Raton, 2008.
[14] Ray, S. C.: Data envelopment analysis: theory and techniques for economics and operations research. Cambridge University Press, New York, 2004.
[15] Staníčková, M., and Melecký, L.: Competitiveness evaluation of Visegrad four countries by CCR input oriented model of data envelopment analysis. *Scientific papers of the University of Pardubice, Series D - Faculty of Economics and Administration* **22**, (2011), 176–188.
[16] Turkan, S., Polat, E., and Gunay, S.: The eficiency analysis of non-life insurance companies active in Turkey. *Zagreb International Review of Economics & Business* **15**, 2 (2012), 1–14.
[17] Viturka, M.: Kvalita podnikatelského prostředí, regionální konkurenceschopnost a strategie regionálního rozvoje České republiky. Grada Publishing, Praha, 2010.

	Year	Capital City of Prague	Central Bohemian	South Bohemian	Pilsen	Karlovy Vary	Ústí	Liberec	Hradec Králové	Pardubice	Vysočina	South Moravian	Olomouc	Zlín	Moravian-Silesian
Number of inhabitants to 31.12.	2011	1 241 664	1 279 345	636 138	571 709	303 165	828 026	438 600	553 856	516 411	511 937	1 166 313	638 638	589 030	1 230 613
	2012	1 246 780	1 291 816	636 611	572 687	301 726	826 764	438 594	552 946	516 440	511 207	1 168 650	637 609	587 693	1 226 602
	2013	1 243 201	1 302 336	636 707	573 469	300 309	825 120	438 609	551 909	515 985	510 209	1 170 078	636 356	586 299	1 221 832
	2014	1 259 079	1 315 299	637 300	575 123	299 293	823 972	438 851	551 590	516 372	509 895	1 172 853	635 711	585 261	1 217 676
	2015	1 267 449	1 326 876	637 834	576 616	297 828	822 826	439 639	551 421	516 149	509 475	1 175 025	634 718	584 676	1 213 311
R&D expenses (in thous. CZK)	2011	22 941	6 350	2 193	3 142	124	843	1 861	1 679	2 472	780	11 192	2 133	2 118	4 924
	2012	24 689	6 677	2 537	3 779	204	1 125	2 860	1 680	2 783	922	14 645	3 558	2 317	4 584
	2013	26 165	9 718	2 534	4 133	115	1 084	2 366	1 890	2 687	1 160	16 185	3 061	2 254	4 500
	2014	29 443	9 879	2 488	4 737	151	1 216	2 614	2 054	2 727	1 502	17 012	3 377	2 749	5 155
	2015	32 999	9 991	2 665	4 607	203	1 097	2 520	1 987	2 650	1 536	17 699	2 983	2 533	5 194
NHDI (in thous. CZK)	2011	324 042	273 227	119 094	112 252	53 413	145 399	79 838	104 346	94 323	96 331	222 140	113 538	107 182	219 743
	2012	323 111	280 494	119 558	114 951	52 663	143 258	80 375	106 304	97 160	96 573	227 734	114 957	108 629	220 074
	2013	324 607	277 187	118 621	116 400	53 288	143 091	80 548	106 313	96 422	97 264	232 222	113 032	107 966	217 907
	2014	334 695	288 382	123 842	118 470	55 095	144 972	83 565	110 491	101 043	100 279	240 632	117 577	112 207	226 407
	2015	344 980	296 468	126 189	124 681	57 310	150 644	86 348	115 094	103 567	103 570	249 889	121 767	117 884	232 690
GDP per employed person (in CZK)	2011	1 544	723	679	735	579	690	648	727	695	689	796	672	713	751
	2012	1 551	721	706	716	595	708	679	729	667	714	808	666	710	753
	2013	1 557	722	717	742	581	687	676	733	671	707	814	689	703	730
	2014	1 611	779	730	797	600	697	694	774	697	730	849	718	781	765
	2015	1 713	813	760	810	602	752	718	805	719	761	876	745	794	790
Employees' average gross monthly salary (in CZK)	2011	34 403	25 605	23 040	24 086	21 568	23 081	23 240	22 697	22 792	22 680	24 518	22 670	22 461	23 909
	2012	35 356	25 923	22 871	24 295	21 663	23 608	23 850	23 371	23 080	23 272	25 153	22 754	22 517	24 340
	2013	35 155	26 302	23 429	24 698	22 333	23 886	24 381	23 639	23 187	23 745	25 587	23 203	23 117	24 397
	2014	35 343	27 046	24 239	26 004	23 008	24 331	25 114	24 348	23 879	24 347	26 079	24 081	23 755	24 667
	2015	36 371	27 997	25 246	27 013	24 119	25 301	26 358	25 192	24 856	25 258	27 051	24 584	24 554	25 475

Table 3 Inputs and outputs

Generalized Dynamic Simulation Model of Rating Alternatives by Agents with Interactions

Radomír Perzina[1], Jaroslav Ramík[2]

Abstract. In this paper we deal with a dynamic system of rating given number of alternatives based on alo-groups where agent-based simulations with interactions among agents are dependent on some parameters. An extension of the previous approach deals with the problem of ranking a finite number of alternatives by pairwise comparison matrices. A dynamic system of rating for a given number of alternatives based on agent-based simulation with interactions among agents, where their PC matrices take elements from an alo-group, is presented. This system is able to simulate various business or social processes, e.g. financial market, products' demand and supply, electors' preferences of political parties, etc. A simple simulation experiment has been performed.

Keywords: dynamic system, agents, ranking alternatives, pairwise comparison matrix on alo-group, simulation.

JEL classification: C61
AMS classification: 90C29

1 Introduction

A significant feature of agent-based modeling relies on the power of computers to explore dynamics out of the reach of pure mathematical methods by means of repetitive interactions between the *agents*, see [3], [2], [9], [7]. In addition, agents may be capable of evolving, allowing unanticipated behaviors to emerge, see [5], [9]. A process of simulation models usually incorporate several basic components that reflect agent behavior for modeling system inputs and outputs and ranking and/or rating given alternatives, see e.g. [4], [12], [9], and [14]. In this paper we deal with a dynamic system of rating given number of alternatives based on alo-groups. In each time moment, a finite set of alternatives is ranked by each agent from the finite set of agents. Then, the total ranking is calculated. In the course of time, the individual agents interact with each other according to the given system of rules. The agent-based simulations with interactions among agents are dependent on some parameters. The system of rating alternatives is able to replicate various processes, e.g. financial market modeling, see [12], [14], auction models, [2], demand and supply models, [4], evaluation of political parties in general elections, evaluation of universities or other public institutions, etc.

The structure of the paper can briefly be described as follows. In Section 2 the basic problem is formulated as well as necessary notions and terminology concerning alo-groups. A detailed agent-based model is described in Section 3. A simple simulation experiment is presented and discussed in Section 4. The conclusion section finalizes the paper.

2 Preliminaries

We shall deal with the following problem: Let $X = \{x_1, x_2, ..., x_n\}$ be a finite set of alternatives ($n > 1$). We consider a set of agents $K = \{1, 2, ..., |K|\}$, e.g. brokers, customers, electors, students etc. Here, by symbol $|K|$ we denote the (finite) number of elements of the set K. The whole group of agents K is composed of various sub-groups of similar agents. In every time moment the agents may interact (communicate) among each others and switch from one sub-group to some other sub-group.

Each agent $k \in K$ makes successively his/her decisions in every discrete time moment $t = 1, 2, ..., T$ by a pairwise comparisons matrix (PC matrix) $A(k, t)$. The aim is to get *global rating*, or, *group rating*, of the given alternatives in each time moment t, using the information given by each agent k in the form of an $n \times n$ *individual pairwise comparisons matrix*

$$A(k, t) = \{a_{ij}(k, t)\}. \tag{1}$$

In [13], given time t each agent k evaluates the pair of alternatives x_i, x_j by a positive real number $a_{ij}(k, t)$, for all i and j. In this paper, we extend this approach to assume that the element $a_{ij}(k, t)$ of the pairwise comparisons

[1]Silesian University in Opava, Faculty of Business Administration in Karviná, University Sq. 1934/3, 733 40 Karviná, Czech Republic, perzina@opf.slu.cz

[2]Silesian University in Opava, Faculty of Business Administration in Karviná, University Sq. 1934/3, 733 40 Karviná, Czech Republic, ramik@opf.slu.cz

matrix $A(k,t)$ belongs to a more general structure then positive real numbers with the operation of multiplication. Particularly, we shall assume that the elements $a_{ij}(k,t)$ belong to an Abelian linearly ordered group G, (shortly *alo-group*) $G \subset \mathbf{R}$, i.e. a subset of the set of real numbers. This approach enables us to apply an agent based approach not only to multiplicative system as described in [9], but also to apply additive systems, fuzzy systems or other systems. Bellow, we shortly remind some necessary terminology associated with alo-groups.

An *abelian group* is a set, G, together with an operation $\odot$ (read: operation *odot*) that combines any two elements $a, b \in G$ to form another element in G denoted by $a \odot b$, see [6]. The symbol $\odot$ is a general placeholder for a concretely given operation. $(G, \odot)$ satisfies the following requirements known as the *abelian group axioms*, particularly: *commutativity*, *associativity*, there exists an *identity element* $e \in G$ and for each element $a \in G$ there exists an element $a^{(-1)} \in G$ called the *inverse element to* a.

The *inverse operation* $\div$ to $\odot$ is defined for all $a, b \in G$ as follows

$$a \div b = a \odot b^{(-1)}. \tag{2}$$

An ordered triple $(G, \odot, \leq)$ is said to be *abelian linearly ordered group*, *alo-group* for short, if $(G, \odot)$ is a group, $\leq$ is a linear order on G, and for all $a, b, c \in G$

$$a \leq b \text{ implies } a \odot c \leq b \odot c. \tag{3}$$

If $\mathcal{G} = (G, \odot, \leq)$ is an alo-group, then $G \subset \mathbf{R}$ is naturally equipped with the order topology induced by $\leq$ and $G \times G$ is equipped with the related product topology. We say that $\mathcal{G}$ is a *continuous alo-group* if $\odot$ is continuous on $G \times G$.

$\mathcal{G} = (G, \odot, \leq)$ is *divisible* if for each positive integer n and each $a \in G$ there exists the (n)-th root of a denoted by $a^{(1/n)}$, i.e. $\left(a^{(1/n)}\right)^{(n)} = a$.

Because of the associative property, the operation $\odot$ can be extended by induction to n-ary operation: $a^{(n)} = a \odot a \odot ... \odot a$. By divisibility, (n)-th power of a can be extended to (c)-th power of a, $a^{(c)}$, where $c \in \mathbf{R}$. It is well known fact, that a continuous alo-group of real numbers is divisible and that it must be an open interval.

We define a "multiplication" operation "$\bullet$" as follows: For all $a \in G, c \in \mathbf{R}$:

$$c \bullet a = a^{(c)}. \tag{4}$$

Then $\mathcal{H} = (G, \odot, \bullet, \leq)$ is the *Riesz space*, with the following properties, see [1]:
- $(G, \odot, \bullet)$ is a vector space over the field $\mathbf{R}$;
- $(G, \leq)$ is a lattice;
- for every $a, b \in G$, $c \geq 0, a \leq b$ implies $c \bullet a \leq c \bullet b$.

Let $\mathcal{G} = (G, \odot, \leq)$ be an alo-group. As every alo-group $\mathcal{G}$ is a lattice ordered group, there exists $\max\{a, b\}$, for each pair $(a, b) \in G \times G$. Nevertheless, a nontrivial alo-group $\mathcal{G} = (G, \odot, \leq)$ has neither the greatest element nor the least element. Then, function $\|.\| : G \to G$ defined for each $a \in G$ by

$$\|a\| = \max\{a, a^{(-1)}\} \tag{5}$$

is called a $\mathcal{G}$*-norm*.

The operation $d : G \times G \to G$ defined by $d(a, b) = \|a \div b\|$ for all $a, b \in G$ is called a $\mathcal{G}$*-distance*. Next, we present the well known examples of alo-groups of real numbers $\mathbf{R}$, see also [6], or, [10].

Example 1: Additive alo-group $\mathcal{R} = (\mathbf{R}, +, \leq)$ is a continuous alo-group with: $e = 0,\ a^{(-1)} = -a,\ a^{(n)} = a + a + ... + a = n.a$.

Example 2: Multiplicative alo-group $\mathcal{R}^+ = (]0, +\infty[, \bullet, \leq)$ is a continuous alo-group with: $e = 1,\ a^{(-1)} = a^{-1} = 1/a,\ a^{(n)} = a^n$. Here, by $\bullet$ we denote the usual operation of multiplication.

Example 3: Fuzzy additive alo-group $\mathcal{R}_a = (\mathbf{R}, +_f, \leq)$, see [10], is a continuous alo-group with: $a +_f b = a + b - 0.5, e = 0.5, a^{(-1)} = 1 - a, a^{(n)} = n.a - (n-1)/2$.

Example 4: Fuzzy multiplication alo-group $]\mathbf{0}, \mathbf{1}[_{\mathbf{m}} = (]0, 1[, \bullet_f, \leq)$, see [6], is a continuous alo-group with: $a \bullet_f b = \frac{ab}{ab+(1-a)(1-b)}, e = 0.5, a^{(-1)} = 1 - a$.

3 Agent-based model

Now, let us return back to our original rating problem. If $a_{ij}(k,t) > e$, then x_i "is better than" x_j. Here, e is the identity element of a real continuous alo-group $\mathcal{G}$. The higher is $a_{ij}(k,t)$, the stronger is agent's evaluation that x_i "is better than" x_j.

On the other hand, if $a_{ij}(k,t) < e$, then x_j "is better then" x_i. The lower is the value of $a_{ij}(k,t)$, the stronger the evaluation that x_j "is better than" x_i.

If $a_{ij}(k,t) = e$, then both alternatives x_j, x_i are evaluated equally.

PC matrix (1) is assumed to be *reciprocal*, which is a natural requirement, see e.g. [6], [10], [11]. Hence, we have

$$a_{ji}(k,t) = a_{ij}(k,t)^{(-1)}, \text{ for all } i,j \in \{1,2,...,n\}, k \in K, t = 1,2,...,T. \tag{6}$$

The global rating of the alternatives $x_1, x_2, ..., x_n$ in time t is associated with the *global priority vector* $w(t) = (w_1(t), ..., w_n(t))$ which is calculated from the *global PC matrix* $A(t) = \{a_{ij}(t)\}$, by aggregation (i.e. by the $\odot$-average over all agents) of individual PC matrices as follows, see [10]:

$$a_{ij}(t) = \left(\bigodot_{k=1}^{|K|} a_{ij}(k,t) \right)^{(\frac{1}{|K|})}, \text{ for all } i,j \in \{1,2,...,n\}, t = 1,2,...,T. \tag{7}$$

The weights w_j of the global priority vector $w(t) = (w_1(t), ..., w_n(t))$ are calculated as row geometric averages of the global PC matrix as follows, [10]:

$$w_i(t) = \kappa(t) \left(\bigodot_{j=1}^{n} a_{ij}(t) \right)^{(\frac{1}{n})}, \text{ for all } i \in \{1,2,...,n\}, t = 1,2,...,T. \tag{8}$$

Here, $\kappa(t)$ is a normalizing factor as the global priority vector should be normalized in every time moment $t = 1,2,...,T$, i.e.

$$\bigodot_{j=1}^{n} w_j(t) = e, \text{ for } t = 1,2,...,T. \tag{9}$$

From (8) and (9) we obtain easily

$$\kappa(t) = \left(\bigodot_{j=1}^{n} \left(\bigodot_{i=1}^{n} a_{ij}(t) \right)^{(\frac{1}{n})} \right)^{(-1)}, \text{ for all } t = 1,2,...,T. \tag{10}$$

The global priority vector is associated with the ranking of alternatives as follows:

$$\text{If } w_i(t) > w_j(t) \text{ then } x_i \succ x_j,$$

where $\succ$ stands for "is better than", [10], [11].

In the course of time, the individual agents interact with each other, see [13], personally, by social networks or otherwise. Particularly, in time $t = 1,2,...,T$, each agent $k \in K$ belongs to exactly one of s agent-types, $s \in S = \{1,2,...,|S|\}$, where $|S| \geq 1$, i.e. he/she belongs to exactly one of disjoint sets $K_1(t), K_2(t), ..., K_{|S|}(t)$ satisfying $K_1(t) \cup K_2(t) \cup ... \cup K_{|S|}(t) = K$, $K_r(t) \cap K_s(t) = \emptyset, r \neq s$. The set of agents $K = \{1,2,...,|K|\}$ is supposed to be constant over time $t = 1,2,...,T$.

Each agent $k \in K$ is characterized by the equation of dynamics of the changes of evaluations of the alternatives as follows:

$$a_{ij}(k,t+1)) = a_{ij}(k,t) \odot [a_{ij}(k,t) \div a_{ij}(k,t-1)]^{(c)} \odot m(k,t), \tag{11}$$

$$a_{ji}(k,t+1)) = a_{ij}(k,t+1)^{(-1)}, \text{for all } i,j \in \{1,2,...,n\}, i < j, \text{ and } k \in K, \tag{12}$$

for all $t = 1,2,...,T$. Here, (12) is the reciprocity condition, moreover, for $i \in \{1,2,...,n\}$ we have $a_{ii}(k,t+1)) = e$, $c \in \mathbf{R}$ is a parameter, $m(k,t)$ is a *random* member (or, *error*) with a given probability distribution, e.g. the normal one $N(e, \sigma(k,t))$.

From (11) it follows that the pairwise comparisons of alternatives given by individual agents in time $t+1$ depend on the previous evaluations in time t and, on a relative increment of evaluations between time t and $t-1$, moreover, a small random value (dependent on a parameter σ) is added to the right hand side of equation (11) in order to express small non-specific effects around identity element e.

In each time $t = 1,2,...,T$, the agents may meet, or interact, at random, and there is a probability that one agent may convince the other agent to follow his/her opinion. In addition, there is also a small probability that an agent changes his/her opinion independently. A key property of this model is that direct interactions between

heterogeneous agents may lead to substantial opinion swings from the set of type $K_r(t)$, into a set of type $K_s(t)$, or, vice versa. Here, $r, s \in S, r \neq s$. As an example, consider $|S| = 2$, $K_1(t)$ is the set of optimistic traders, $K_2(t)$ is the set of pessimistic traders.

For each sub-group of agents $K_s(t)$ of type $s \in S, t = 1, 2, ..., T$, we calculate the *group PC matrix* $A^s(t) = \{a^s_{ij}(t)\}$ as $\odot$-mean as follows:

$$a^s_{ij}(t) = \left(\bigodot_{k \in K_s(t)} a_{ij}(k,t) \right)^{(\frac{1}{|K_s(t)|})}, \text{ for all } i, j \in \{1, 2, ..., n\}. \tag{13}$$

Here again, by $|K_s(t)|$ we denote the number of elements (i.e. agents) in the set $K_s(t)$.

Given $t \in \{1, 2, ..., T\}$, $s \in S$. For each $k \in K$ we compute the "distance" $d(k, s, t)$ of PC matrix $A(k, t)$ from $A^s(t)$, i.e.

$$d(k, s, t) = \|\{a_{ij}(k, t) \div a^s_{ij}(t)\}\|. \tag{14}$$

Here, $\|...\|$ is a matrix norm, e.g. for an $n \times n$ matrix $A = \{a_{ij}\}$, we use the norm, see e.g. [10]:

$$\|A\| = \max\{\max\{a_{ij}, a^{(-1)}_{ij}\} | i, j = 1, ..., n\}. \tag{15}$$

For $k \in K_s(t)$, denote

$$d(k, s^*, t) = \min\{\ d(k, s, t) | s \in \{1, 2, ..., S\}\}. \tag{16}$$

If $d(k, s, t) > d(k, s^*, t)$, then in time $t + 1$ the agent k switches from type s to type s^*, i.e.

$$K_s(t + 1) = K_s(t) - \{k\}, K_{s^*}(t + 1) = K_{s^*}(t) \cup \{k\}. \tag{17}$$

Therefore, the number of agents will change as follows:

$$|K_s(t + 1)| = |K_s(t)| - 1, \text{ and } |K_{s^*}(t + 1)| = |K_{s^*}(t)| + 1. \tag{18}$$

If $d(k, s, t) \leq d(k, s^*, t)$, then, in time $t + 1$, the agent k does not switch into any other type group. In this way we perform all switches of agents in K, i.e. we obtain new sets $K_s(t + 1), s \in \{1, 2, ..., S\}$ in time moment $t + 1$.

4 Simulation experiment

Our problem formulated in Section 2 and Section 3 will be illustrated on a simulation experiment. Let $X = \{x_1, x_2, ..., x_n\}$ be a set of n alternatives. Here, we assume the additive alo-group $\mathcal{R} = (\mathbf{R}, +, \leq)$ from Example 1, i.e. $\odot = +$. Then, formulas (7) - (11) can be then reformulated as follows:

$$a_{ij}(t) = \frac{1}{|K|} \sum_{k=1}^{|K|} a_{ij}(k, t), \text{ for all } i, j \in \{1, 2, ..., n\}, t = 1, 2, ..., T; \tag{19}$$

$$w_i(t) = \kappa(t) + \frac{1}{n} \sum_{j=1}^{n} a_{ij}(t), \text{ for all } i \in \{1, 2, ..., n\}, t = 1, 2, ..., T; \tag{20}$$

$$\sum_{j=1}^{n} w_j(t) = 0, \text{ for all } t = 1, 2, ..., T; \tag{21}$$

$$\kappa(t) = -\frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij}(t), \text{ for all } t = 1, 2, ..., T; \tag{22}$$

Dynamic changes of the individual PC matrix (11) are reformulated by the following equations:

$$a_{ij}(k, t + 1)) = a_{ij}(k, t) + c.[a_{ij}(k, t) - a_{ij}(k, t - 1)] + m(k, t), \tag{23}$$

$$a_{ji}(k, t + 1)) = -a_{ij}(k, t + 1), \text{ for all } i, j \in \{1, 2, ..., n\}, i < j, \text{ and } k \in K,$$

where $t = 1, 2, ..., T$. Moreover, for $i \in \{1, 2, ..., n\}$ we set the diagonal elements of the PC matrix as identity elements, i.e. $a_{ii}(k, t+1)) = 0$. Here, $m(k, t)$ is a random member (error) with normal distribution $N(0, \sigma(k, t))$, where $\sigma(k, t) = \sigma$, and c is a parameter.

Now, we consider $n = 4$, the set of agents $K = \{1, 2, ..., 15\}$, and $T = 100$, where each agent $k \in K$ makes successively his/her decisions in every discrete time moment $t = 1, 2, ..., 100$ by a PC matrix $A(k, t)$ with the dynamic equation (23). The aim is to investigate global rating of the alternatives in each time moment t, particularly in the final time $t = 100$, which can be interpreted as a time of prediction of ranking of alternatives. To this aim we use information given by each agent k in the form of a 4×4 individual PC matrix. The computer program for this simulation experiment was prepared in $C\#$.

The simulation starts with 4×4 PC matrix for $k \in K$:

$$A(k,0) = \{a_{ij}(k,0)\} = \begin{pmatrix} 0 & 2 & 4 & 6 \\ -2 & 0 & 2 & 4 \\ -4 & -2 & 0 & 2 \\ -6 & -4 & -2 & 0 \end{pmatrix} = A(k,1) = \{a_{ij}(k,1)\}. \tag{24}$$

Here, we use the well known additive scale $-9, -8, ..., -1, 0, 1, 2, ..., 8, 9$, see [8], with the following interpretation of comparisons of pairs x versus y:

0 ... x is equally important to y,
3 ... x is slightly more important to y,
5 ... x is strongly more important to y,
7 ... x is very strongly more important to y,
9 ... x is absolutely more important to y.

Values 1, 2, 4, 6 and 8 are intermediate values, e.g. 4 means that x is between slightly more important and strongly more important to y. The negative values denote the reciprocal importances. The priority vector $w(1) = (3; 1; -1; -3)$, corresponding to $A(k, 0)$, generates the rating of alternatives: $x_1 > x_2 > x_3 > x_4$.

Each agent belongs to exactly one of $|S| = 2$ agent-types, i.e. in time t, two disjoint sub-groups $K_1(t), K_2(t)$ satisfying $K_1(t) \cup K_2(t) = K$, $K_1(t) \cap K_2(t) = \emptyset$. The set of all agents $K = \{1, 2, ..., 15\}$, is constant in time $t = 1, 2, ..., 100$. The initial matrices $A(k, 0), A(k, 1)$, $k = 1, 2, 3, ..., 15$, are given by (24). The random error m has the normal distribution $N(0, \sigma)$. Here, in the time moment $t = 1$, the agents belong to 2 initial types.

Again, the dynamics of the changes of evaluations of the alternatives by agent $k \in K_s(t)$ is given by formula (23) with the values of $\sigma(k, t) = \sigma$ for all k and t, e.g. $\sigma = 1.0$, or, $\sigma = 3.0$. Moreover, c is a parameter with the value less than 1, e.g. $c = 0.1$, or, $c = 0.3$. For computing distances of an agent's PC matrix to the global PC matrix, we use norm (15). Changes of number of individual agents of a given type are given by formulas (17) and (18). For an illustration, the results of simulation computations for various values of parameter σ and c in the form of time series of the global weights of the alternatives are depicted.

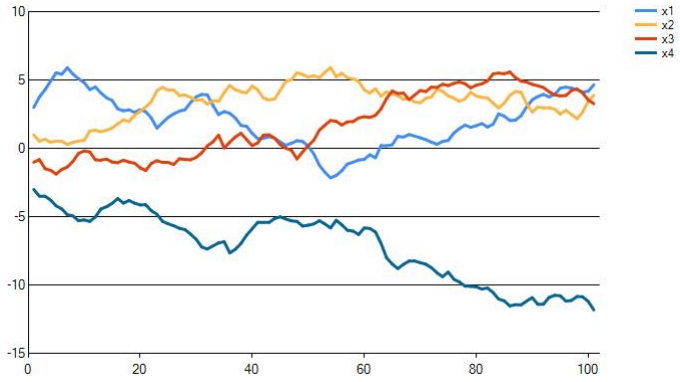


Fig. 1: Global rating of 4 alternatives: $\sigma = 3.0, c = 0.3$

In Fig.1, we can see the time series of the global weights $w_i(t), i = 1, 2, 3, 4$, of alternatives x_i in the course of time t, $t = 1, 2, ..., 100$. Here, we consider a relatively small error m, i.e. small values of the parameters $\sigma = 3.0, c = 0.3$ Then, in the final time moment $t = 100$ the vector of weights will not change much, particularly with the same ranking of alternatives: $x_1 > x_2 > x_3 > x_4$. In Fig.2, the number of agents of both types are depicted depending on time t.

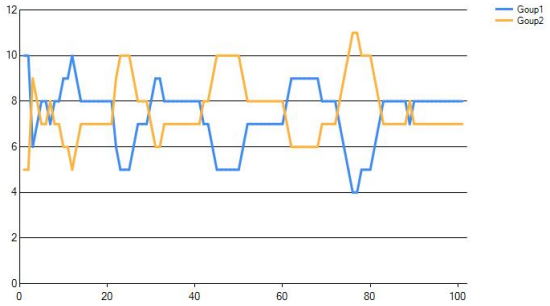


Fig. 2: Number of agents of 2 types, $\sigma = 3.0, c = 0.3$

5 Conclusion

In this paper we present an extension of the approach from [9]. A dynamic system of rating for a given number of alternatives based on agent-based simulation with interactions among agents, where their PC matrices take elements from a general alo-group, is presented. This system is able to simulate various business or social processes, e.g. financial market, products' demand and supply, electors' preferences of political parties, etc. A simple simulation experiment has been performed. It turns out that for small values of the parameters σ and c the global weights of alternatives remain approximately constant in the course of time. When the values of the parameters are increased then in the final time moment the vector of weights changes essentially as well as the rating of the alternatives. More simulation calculations are to be done. In the future, it will be also interesting to compare the results of various types of simulation models, e.g. additive, multiplicative, or fuzzy ones.

Acknowledgements

This paper was supported by the Ministry of Education, Youth and Sports Czech Republic within the Institutional Support for Long-term Development of a Research Organization in 2017.

References

[1] Aliprantis, C. D., Burkinshaw, O.: *Locally solid Riesz spaces with applications to economics.* No. 105. American Mathematical Soc., 2003.

[2] Anthony, P., Jennings, N.R.: Evolving bidding strategies for multiple auctions. In: *Proc. of 15th European Conference on Artificial Intelligence*, Netherlands, 182–187, (2002).

[3] Axelrod, R.: *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton Univ. Press, Princeton, NJ, 1997.

[4] Barnett, M.: Modeling and simulation in business process management. http://news.bptrends.com/publicationfiles, (2013).

[5] Bonabeau, E.: Agent-based modeling: Methods and techniques for simulating human systems. *PNAS*, **99** (2002) 7280–7287.

[6] Cavallo, B., D'Apuzzo, L.: A general unified framework for pairwise comparison matrices in multicriteria methods. *International Journal of Intelligent Systems* **24** 4 (2009) 377–398.

[7] Dumas, M., La Rosa, M., Mendling, J., Reijers, H.: *Fundamentals of Business Process Management*. Springer-Verlag Berlin Heidelberg, 2013.

[8] Mittelhammer, R.C., Judge, G.G., Miller, D.J.: *Econometric Foundations*. Cambridge University Press, 2000.

[9] Perzina, R., and Ramik, J.: Dynamic system of rating alternatives by agents with interactions. In: *Smart Innovation Systems and Technologies* Springer, **58** (2016), 177–185.

[10] Ramik, J.: Pairwise comparison matrix with fuzzy elements on alo-group. *Information Sciences* **297** (2015) 236–253.

[11] Saaty, T.L.: *Multicriteria decision making - the Analytical Hierarchy Process*. Vol. I., RWS Publications, Pittsburgh, 1991.

[12] Sperka, R., Spisak, M.: Transaction costs influence on the stability of financial market: agent-based simulation. *Journal of Business Economics and Management* **14** 1 (2013) 1–12.

[13] Walsh, W.E. et al.: Analyzing complex strategic interactions in multi-agent games. In: *Proc. of AAAI Workshop on Game-Theoretic and Decision-Theoretic Agents*, 1–11, (2002).

[14] Wooldridge, M.: *MultiAgent Systems: An Introduction*. 2nd edition. Chichester: John Wiley and Sons, 2009.

Transit coordination in bus-railway networks

Štefan Peško[1], Tomáš Majer[2]

Abstract.
We present an optimization model for modification of transit timetables that maximizes transfers. We attempt to find an optimal timetable by shifting the existing timetable and/or adding holding time to the timetable to minimize delays associated with transfers from a feeder route to a receiving route. The model of the Integer Linear Programming (ILP) is developed to estimate optimum schedule shifts in timetables. The developed model is evaluated through a real bus-railway instance of slovak city Žilina. The waiting times of the passengers transfer for original and modified schedules are calculated. The results show that the model effectively reduces total transfer and waiting times through presented modification of the bus schedule.

Keywords: Transit coordination, Waiting time, Timetables, ILP

JEL classification: C02, C61, C65, C68, J33
AMS classification: 90C10, 90C15

1 Introduction

In bus-railway transit networks with several bus lines going through train station the transfer time is a significant component of overall travel time from passengers perspective. Thus, efficient transfers are very important for increasing attractiveness and productivity of transit service. Coordination of transit schedules can reduce transfer time significantly.

Local researches dealing with coordinating multiple bus lines focused more heavily on scheduling rather than evaluating behavioral choices of passengers as we can see in Gabrisova, Kozel [3], Janacek, Kohani [5] and Paluch [6]. Timetables among intersecting transfers from a feeder route to a receiving route between rail station and bus line stops may be coordinated via shifting departure times of trains or departure times of buses.

Ciaffi, Cipriani, Petrelli [2] deals with the transit network design problem related to problem when feeder routes are defined as transit services for the connection of local areas with the stops of the main transit network usually railway or underground station. Their two-phase procedure is based on solving traveling salesman problem for connecting the highest demand node pairs in the area with the stop of main transit network and k-shortest path algorithm for feasible set aims at developing feeder routes.

Studies with joint optimization of a rail transit route and bus routes in a transit corridor in many large cities of China to maximize rail ridership and minimize total passenger travel time were developed. Yang, Xiaonian, Baoqing Dehui [8] proposed method for classifying bus routes and developed multi-objective model for designing an integrated rail transit and bus network based on genetic algorithm.

Dynamic holding strategy for transfer coordination based on control theory was studied in Anderson, Daganzo [1] where maximum holding time is a function of real-time estimates of bus arrivals and passengers and the uncertainty in these estimates.

We present an optimization model for modification of transit timetables that minimize transfer-related times. In this model we attempt to find an optimal timetable by shifting the existing timetable and/or adding holding time to the timetable to minimize delays associated with transfers from a feeder route to a receiving route. The model of the Integer Linear Programming (ILP) is developed to estimate optimum schedule shifts in timetables and evaluated through a real bus-railway instance from slovak city Žilina.

2 Mathematical formulation

The essential element of the bus-railway transit coordination are bus trips and railway timetable. A unit of time of all data in this paper is minute.

Given rail timetable by the set $\mathcal{T}^R = \{t_k^R, k \in M\}$, the rail trip is triple $t_k^R = \left(r(t_k^R), a(t_k^R), d(t_k^R)\right)$ where

[1]University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, Žilina, slovakia , Stefan.Pesko@fri.uniza.sk
[2]University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, Žilina, slovakia, Tomas.Majer@fri.uniza.sk

- M - number of trains,
- $r(t_k^R)$ - type of the k-th train,
- $a(t_k^R)$ - arrival time of the k-th train,
- $d(t_k^R)$ - departure time of the k-th train.

For a more elaborate formulation of the model is set $\mathcal{T}^R$ ordered by arrival time i.e.

$$a(t_1^R) \leq a(t_2^R) \leq \cdots \leq a(t_{|M|}^R).$$

Given set of bus trips $\mathcal{T}^B = \{t_i^B, i \in N, \}$, the bus trip is septuple

$$t_i^B = \big(l(t_i^B), dp(t_i^B), dt(t_i^B), ap(t_i^B), at(t_i^B), dts(t_i^B, wt(t_i^B))\big),$$

where

- N - number of trips,
- $l(t_i^B)$ - number of link of the i-th bus trip,
- $dp(t_i^B)$ - departure place of the i-th trip
- $dt(t_i^B)$ - departure time of the i-th trip,
- $ap(t_i^B)$ - arrival place the i-th trip,
- $at(t_i^B)$ - arrival time of the i-th trip,
- $dts(t_i^B)$ - departure time of rail stop of the i-th trip,
- $wt(t_i^B)$ - walking distance time between rail stop and rail station of the i-th trip.

Analogical, but more detailed, we have ordered the set of bus trips $\mathcal{T}^B$. Let's have two bus trips t_1^B and t_2^B). We will say the trip t_1^B precedes the trip t_2^B

$$t_1^B \prec t_2^B \text{ iff } dt(t_1^B) \leq dt(t_2^B) \text{ and } at(t_1^B) + \tau \leq dt(t_2^B).$$

Running board of a bus is a sequence of trips $t_1^B, t_2^B, \ldots, t_r^B$ such that

$$t_1^B \prec t_2^B \prec \cdots \prec t_r^B.$$

In the basic model we assume that we know the following parameters:

- w_{min}^R and w_{max}^R - minimum and maximum waiting time on rail station,
- w_{min}^B and w_{max}^B - minimum and maximum waiting time on bus rail stops,
- δ - maximum time feasible $\pm$ shift of bus trips,
- τ - time of a minimal safe-up of the departure/arrival places,
- p - number of buses of bus schedule.

The objective function is the number of events when the train passengers have bus connection or the bus passengers have a train connection. We can formally define our optimization problem as follows

$$(s_1^*, s_2^*, \ldots) = argmax \sum_{i \in N: -\delta \leq s_i \leq \delta} c(t_i^B, s_i), \tag{1}$$

where

- s_i - feasible time shift of bus trip t_i^B in a running board,
- $c(t_i^B, s_i)$ - number of train/bus trips connected from bus/train trips.

Note that in definitions of train trip and bus trip are omitted place of the rail station and bus stops. It is correct because we use walking times distance between one rail station and the earliest bus stop of links.

3 The ILP model

We will use following denotation of index sets:

- $N = \{1, 2, \ldots, n\}$ - order of the bus trips,
- $S_\delta = \big\{-\delta, -\delta + 1, \ldots, 0, \ldots, \delta - 1, \delta\big\}$ - possible time shifts of any bus trips.

Let us define bivalent variables y_{is} for $i \in N, s \in S_\delta$, by $y_{is} = 1$ if bus trip t_i^B has shifted departure and arrival times i.e $dt(t_i^B) + s$ and $at(t_i^B) + s$. Let us define a decision variable $x_{ij}, i < j, i \in N, j \in N$ with values defined by

$$x_{ij} = \begin{cases} 1 & \text{if shifted } t_i^B \prec \text{ shifted } t_j^B \text{ in a running board} \\ 0 & \text{otherwise .} \end{cases}$$

Now we can formulate our coordination problem as following zero-one linear programming problem (ILP):

$$\sum_{i \in N} \sum_{s \in S_\delta} c(t_i^B, s) \cdot y_{is} \rightarrow maximum \tag{2}$$

$$\sum_{s \in S_\delta} y_{is} = 1 \qquad \forall i \in N, \tag{3}$$

$$\sum_{i \in N: i<j} x_{ij} \leq 1 \qquad \forall j \in N, \tag{4}$$

$$\sum_{j \in N: i<j} x_{ij} \leq 1 \qquad \forall i \in N, \tag{5}$$

$$\sum_{(i,j) \in N \times N: i<j} x_{ij} = |N| - p \tag{6}$$

$$at(t_i^B) + \sum_{s \in S_\delta} s \cdot y_{is} + (\tau + M) \cdot x_{ij} \leq dt(t_j^B) + \sum_{s \in S_\delta} s \cdot y_{js} + M \qquad \forall (i,j) \in N \times N : i < j, \tag{7}$$

$$x_{ij} \in \{0,1\} \qquad \forall (i,j) \in N \times N : i < j, \tag{8}$$

$$y_{is} \in \{0,1\} \qquad \forall i \in N, \forall s \in S_\delta. \tag{9}$$

Objective function (2) counts total number of connection between train station and bus stops. Constraints (3) select one time shift of the bus trip. Constraints (4) and (5) define matching of bus trips in running boards of bus schedule. Constraints (6) ensure given number of buses p of bus schedule. The admissibility of shifted bus trips guarantees the condition (7), where M is a big penalization constant. Constrains (8) and (9) are obligatory.

4 Real bus-railway instance

Developed ILP model was evaluated via real bus-railway instance from slovak city Žilina during working day and weekend. The transit bus network of city [7] is depicted on figure 1. For our experiment we were limited to 98 fast trains which stop in rail station (Žel. stanica) during working days and weekend.

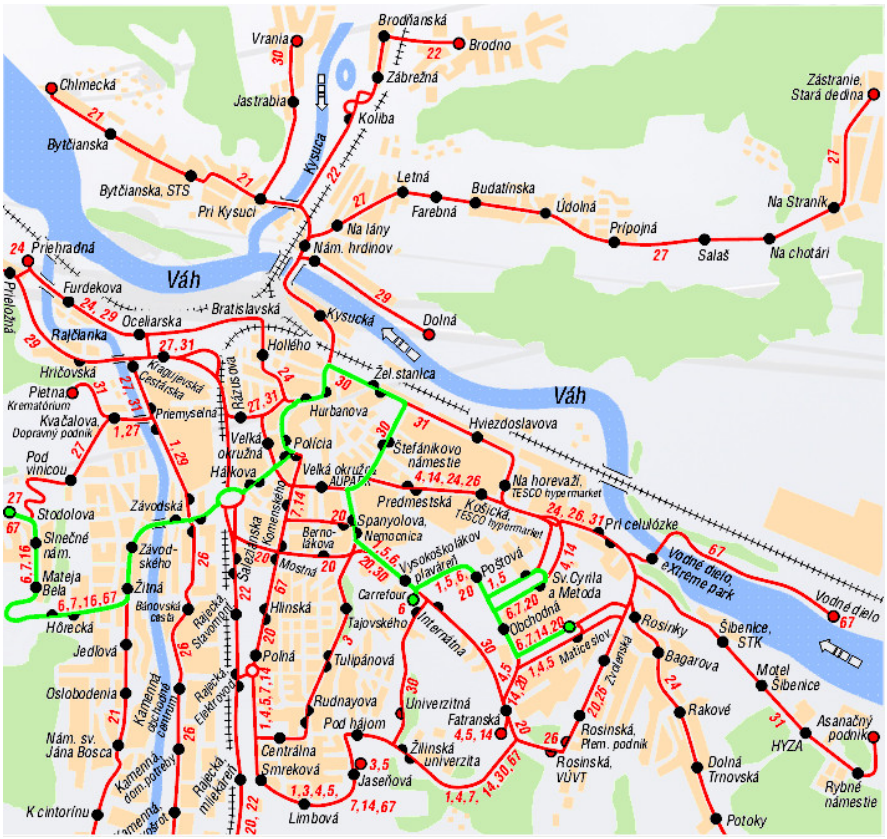


Figure 1 Transit network map of Žilina with bold link 6.

The public transport of city is represented by selected auto-bus a trolleybus lines [7] with stops near the railway station, see tab.1.

Lines	vehicle	weekend trips	workdays trips	two-way
1	troleybus	0	11	no
3	troleybus	47	67	no
4	troleybus	49	70	no
6	troleybus	75	82	yes
14	troleybus	49	72	no
16	troleybus	0	26	no
22	autobus	36	81	yes
24	autobus	31	65	yes
26	autobus	21	46	yes
29	autobus	15	22	yes
30	autobus	10	13	no
31	autobus	4	10	no
50	troleybus	8	8	yes
$\sum$	–	326	573	–

Table 1 Basic characteristics of selected bus-lines of Žilina

In our model we do not distinguish whether a vehicle is a trolleybus or autobus and so we simply talk about the bus. The bus-lines are one-way (circular) or two-way. The stops of lines are different from one another and so constraints (7) guarantee that running boards of buses on lines can not change the vehicle type.

The counts and the waiting times of the passengers transfer for original schedules are presented for weekend in tables 2 and for workdays in table 3.

Bus lines in weekend	3	6	4	14	29	22	30	24	31	26	50
Counts of transfers of train-bus	30	42	16	16	7	25	5	18	4	1	5
Mean waiting time of train-bus	6.4	5.3	6.8	6.3	4.3	5.0	8.0	6.2	5.2	7.0	6.0
Counts of transfers of bus-train	64	89	65	62	23	46	12	38	7	3	3
Mean waiting time of bus-train	20.5	21.3	19.5	19.7	21.1	19.5	21.8	21.7	19.4	23.7	20.7

Table 2 Original counts of passenger transfers and mean waiting times for weekend

Bus lines during workday	1	3	4	6	14	16	29	22	30	24	31	26	50
Counts of transfers of train-bus	3	40	29	45	37	12	9	44	6	29	7	25	5
Mean waiting time of train-bus	6.3	6.4	6.8	5.4	6.1	5.8	5.3	5.5	7.3	5.9	6.9	5.0	6.0
Counts of transfers of bus-train	7	89	93	100	88	39	33	102	11	76	11	74	3
Mean waiting time of bus-train	18.9	20.1	20.0	21.2	20.4	20.3	19.4	20.2	22.2	20.0	21.6	20.4	20.7

Table 3 Original number of passenger transfers and mean waiting times for workday

5 Computation experiment

We did computational experiments on PC with Intel Xeon 32 cores, 2,4 GHz, 256 GB RAM via Gurobi solver [4]. We used following parameters for our computation experiment.

- $w^R_{min} = 10$ and $w^R_{max} = 30$ - minimum and maximum waiting time on rail station,
- $w^B_{min} = 2$ and $w^B_{max} = 10$ - minimum and maximum waiting time on bus rail stops,
- $\delta = 1, 2$ - maximum time feasible $\pm$ shift [minutes] of bus trips,
- $\tau = 10$ - time of a minimal safe-up of the departure/arrival places

- $p_{week} = 18$ and $w_{work} = 44$ - number of busses during weekend and workdays.

The computer time of all instabilities was surprisingly small, it did not exceed 30 seconds, although our ILP is a problem of bivalent programming.

5.1 Weekend

First we calculated the mean waiting times from the computed shifts $\delta = 1$ and $\delta = 2$ of the departure times with 326 bus's trips. The count and the waiting times of the passengers transfer for and modified bus schedules are presented in tables 4 and 5.

Bus lines in weekend	3	6	4	14	29	22	30	24	31	26	50
Counts of transfers of train-bus	34	45	22	24	8	26	5	22	3	1	6
Mean waiting time of train-bus	6.2	5.0	5.8	7.3	4.9	4.8	8.0	6.7	6.7	8.0	5.5
Counts of transfers of bus-train	64	112	69	73	23	50	12	40	10	3	3
Mean waiting time of bus-train	20.3	20.9	20.1	19.2	21.0	19.6	21.8	21.6	18.8	23.0	20.0

Table 4 Weekend's instance with possible shifted of bus trips by ± 1 minute.

Bus lines during weekend	3	6	4	14	29	22	30	24	31	26	50
Counts of transfers of train-bus	37	52	28	32	8	26	8	27	4	1	6
Mean waiting time of train-bus	6.0	5.6	7.0	7.1	3.9	4.7	6.9	6.1	5.8	8.0	6.2
Counts of transfers of bus-train	75	112	79	76	24	50	12	44	9	3	4
Mean waiting time of bus-train	19.6	21.0	19.7	19.5	20.4	19.7	21.9	20.4	19.6	22.7	21.5

Table 5 Weekend's instance with possible shifted of bus trips by ± 2 minute.

Basic characteristics of instances that were computed in software R are depicted on figure 2 The results con-

```
ConnTrainBus0   WaitTrainBus0   ConnBusTrain0   WaitBusTrain0
Min.   : 1.00   Min.   :4.300   Min.   : 3.00   Min.   :19.40
1st Qu.: 5.00   1st Qu.:5.250   1st Qu.: 9.50   1st Qu.:19.60
Median :16.00   Median :6.200   Median :38.00   Median :20.70
Mean   :15.36   Mean   :6.045   Mean   :37.45   Mean   :20.81
3rd Qu.:21.50   3rd Qu.:6.600   3rd Qu.:63.00   3rd Qu.:21.50
Max.   :42.00   Max.   :8.000   Max.   :89.00   Max.   :23.70

ConnTrainBus1   WaitTrainBus1   ConnBusTrain1    WaitBusTrain1
Min.   : 1.00   Min.   :4.800   Min.   :  3.00   Min.   :18.80
1st Qu.: 5.50   1st Qu.:5.250   1st Qu.: 11.00   1st Qu.:19.80
Median :22.00   Median :6.200   Median : 40.00   Median :20.30
Mean   :17.82   Mean   :6.264   Mean   : 41.73   Mean   :20.57
3rd Qu.:25.00   3rd Qu.:7.000   3rd Qu.: 66.50   3rd Qu.:21.30
Max.   :45.00   Max.   :8.000   Max.   :112.00   Max.   :23.00

ConnTrainBus2   WaitTrainBus2   ConnBusTrain2    WaitBusTrain2
Min.   : 1.00   Min.   :3.900   Min.   :  3.00   Min.   :19.50
1st Qu.: 7.00   1st Qu.:5.700   1st Qu.: 10.50   1st Qu.:19.65
Median :26.00   Median :6.100   Median : 44.00   Median :20.40
Mean   :20.91   Mean   :6.118   Mean   : 44.36   Mean   :20.55
3rd Qu.:30.00   3rd Qu.:6.950   3rd Qu.: 75.50   3rd Qu.:21.25
Max.   :53.00   Max.   :8.000   Max.   :112.00   Max.   :22.70
```

Figure 2 Summary of basic characteristics of weekend.

firmed the expectation that maximizing the number of connections minimizes the mean waiting time for passengers.

5.2 Workday

Similarly we calculated the mean waiting time from the computed shifts $\delta = 1$ and $\delta = 2$ of the departure time with 573 bus's trips. The count and the waiting time of the passengers transfer for and modified bus schedules are presented in tables 6 and 7.

Bus lines during workday	1	3	4	6	14	16	29	22	30	24	31	26	50
Counts of transfers train-bus	4	45	36	48	46	15	14	47	7	34	9	30	6
Mean waiting time of train-bus	5.8	6.4	6.3	5.4	6.1	6.1	6.3	6.0	6.4	6.4	6.0	5.5	5.7
Counts of transfers of bus-train	8	92	96	120	99	41	34	114	12	85	11	79	3
Mean waiting time bus-train	20.1	20.2	20.2	20.9	20.9	19.9	19.9	20.4	21.2	20.2	20.9	21.0	20.0

Table 6 Workday's instance with possible shifted of bus trips by ± 1 minutes.

Bus lines during workday	1	3	4	6	14	16	29	22	30	24	31	26	50
Counts of transfers train-bus	5	48	42	57	52	17	16	54	9	40	10	31	6
Mean waiting time of train-bus	5.4	6.1	6.8	5.7	6.0	5.3	6.6	5.6	6.6	5.8	6.0	5.5	6.2
Counts of transfers of bus-train	9	104	106	122	108	44	36	118	13	90	12	81	4
Mean waiting time of bus-train	18.9	19.6	20.0	20.8	19.7	19.1	20.1	20.0	21.9	20.3	21.5	21.0	21.5

Table 7 Workday's instance with possible shifted of bus trips by ± 2 minutes.

Also, the results in the bigger instance have confirmed the expectation that maximizing the number of connections minimizes the mean waiting time for passengers.

```
ConnTrainBus0   WaitTrainBus0   ConnBusTrain0    WaitBusTrain0
Min.   : 3.00   Min.   :5.000   Min.   :  3.00   Min.   :18.90
1st Qu.: 7.00   1st Qu.:5.500   1st Qu.: 11.00   1st Qu.:20.00
Median :25.00   Median :6.000   Median : 74.00   Median :20.30
Mean   :22.38   Mean   :6.054   Mean   : 55.85   Mean   :20.42
3rd Qu.:37.00   3rd Qu.:6.400   3rd Qu.: 89.00   3rd Qu.:20.70
Max.   :45.00   Max.   :7.300   Max.   :102.00   Max.   :22.20

ConnTrainBus1   WaitTrainBus1   ConnBusTrain1    WaitBusTrain1
Min.   : 4.00   Min.   :5.400   Min.   :  3.00   Min.   :19.90
1st Qu.: 9.00   1st Qu.:5.800   1st Qu.: 12.00   1st Qu.:20.10
Median :30.00   Median :6.100   Median : 79.00   Median :20.20
Mean   :26.23   Mean   :6.031   Mean   : 61.08   Mean   :20.45
3rd Qu.:45.00   3rd Qu.:6.300   3rd Qu.: 96.00   3rd Qu.:20.90
Max.   :48.00   Max.   :6.400   Max.   :120.00   Max.   :21.20

ConnTrainBus2   WaitTrainBus2   ConnBusTrain2    WaitBusTrain2
Min.   : 5.00   Min.   :5.300   Min.   :  4.00   Min.   :18.90
1st Qu.:10.00   1st Qu.:5.600   1st Qu.: 13.00   1st Qu.:19.70
Median :31.00   Median :6.000   Median : 81.00   Median :20.10
Mean   :29.77   Mean   :5.969   Mean   : 65.15   Mean   :20.34
3rd Qu.:48.00   3rd Qu.:6.200   3rd Qu.:106.00   3rd Qu.:21.00
Max.   :57.00   Max.   :6.800   Max.   :122.00   Max.   :21.90
```

Figure 3 Summary of basic characteristics of workday.

6 Conclusion and future work

Computation experiment on small but real instance has shown that the presented model for transit coordination in bus-railway network with one rail station can be solved exactly. Unlike well-known models, it offers an acceptable running board of buses on coordinated lines for small shifts of bus trips.

In the further research, we want to focus on the following issues:

- In the case of city Žilina, solve problem with all trains stopped in station, not only fast train.
- Verify the model for big instances.
- To work in the model with some time-dependent parameters, e.g. a range of waiting time on rail station or bus stops.
- To use also uncertainties of data that are achievable from transport companies.

Acknowledgements

The research was supported by the research grants APVV-14-0658 Optimization of urban and regional public personal transport.

References

[1] Anderson, P., Daganzo, C.: *Coordinating Transit Transfers in Real Time*, Final Report UCCONNECT 2016 TO 021 - 65A0529, University of California Berkeley, 2016.

[2] Ciaffi, F., Cipriani, E., Petrelli, M.: Feeder bus betwork besign broblem: a new metaheuristic procedure and real size applications,*Procedia* - Social and Behavioral Sciences, Vol., **54** (2012), 798–807.

[3] Gabrisova, L., Kozel, P.: Coordianation of bus departures by mathenatical programming, *Communications* scientific letters of the University of Zilina, Vol. **1**, (2013), 74–79.

[4] Gurobi Optimization, Inc. Gurobi Optimizer Reference Manual, (2015), Available at: http://www.gurobi.com

[5] Janacek, J. Kohani, M.: Waiting time optimisation with IP-solver, *Communications* - scientific letters of the University of Zilina, Vol. **12** No. 3A, (2010), 36–41.

[6] Paluch, S.: Trip coordination in municipal passenger transport, *Transport problems*, Vol. **9**, iss.2, (2014), 111–117.

[7] Time table, Transport enterprise of the city of Žilina, (2017) Available at: http://www.dpmz.sk/en/time-table/.

[8] Yang S., Xiaonian S., Baoqing L., Dehui G.: Joint optimization of a rail transit route and bus routes in a transit corridor, *Procedia - Social and Behavioral Sciences*, 96, (2013), 1218–1226.

Expected return rate determined as oriented fuzzy number

Krzysztof Piasecki[1]

Abstract. The starting point for our discussion is the present value (PV) determined by means of positive L-R fuzzy number. In this paper the information described by the so-determined PV is supplemented with a subjective forecast of the sense trend of observed current market price. This forecast is implemented in the proposed PV model, as the orientation of fuzzy number. The assumption of market price increase is described as a positive orientation of ordered fuzzy number. In analogous way, the assumption of market price decrease is described as the negative orientation of ordered fuzzy number. In this way, PV is presented as ordered fuzzy number [2]. So specified PV is used for determine the return rate which is defined as any decreasing function of PV and increasing function of future value (FV). With the obvious assumption that the FV is a random variable, determined return rate is described as ordered fuzzy random variable. At the end it is shown that the expected return rate is determined as ordered fuzzy number. The orientation of expected return rate is opposite to orientation of the PV defining it.

Keywords: present value, return rate, imprecision, ordered fuzzy number.

JEL Classification: C44, C02, G10
AMS Classification: 03E72, 91G10

1 Introduction

In the most general way, return rate is defined as increasing function of future value (FV) and decreasing function of present value (PV). FV is usually defined as random variable. In [6] return rate is determined for the case when PV is given as fuzzy number. An example of so-defined PV is Behavioural PV (BPV) dependent on selected behavioral factors described in [14]. In [11] the information described by BPV was supplemented with a subjective forecast of orientation of the market price trend. This forecast was implemented in the model BPV as an orientation of ordered fuzzy number [9]. In [11] oriented BPV was applied for determining simple return rate.

The main goal of this paper is to determine generalized return rate for the case when PV is described by ordered fuzzy number. For so-defined return rate will be determined expected return rate.

2 The essence of the notion of ordered fuzzy number

The family of all fuzzy subsets in the real line $\mathbb{R}$ is denoted by the symbol $\mathcal{F}(\mathbb{X})$. An imprecise number is a family of values in which each considered value belongs to it in a varying degree. A commonly accepted model of imprecise number is the fuzzy number (FN), defined in [3]. We will to consider the following type of FN.

Definition 1: [4] For any quadruple $(a,b,c,d) \in \mathbb{R}^4$ such that $a \le b \le c \le d$, the LR-type FN $S(a,b,c,d) \in \mathcal{F}(\mathbb{X})$ is described by its membership function $\mu_S \in [0;1]^{\mathbb{R}}$ defined as follows

$$\mu_S(x) = \begin{cases} 0 & x < a \\ L_S(x) & a \le x \le b \\ 1 & b \le x \le c \\ R_S(x) & c \le x \le d \\ 0 & d < x \end{cases}, \tag{1}$$

where:

- the left reference function $L_S \in [0,1]^{[a,b]}$ is continuous from above and increasing one;
- the right reference function $R_S \in [0,1]^{[c,d]}$ is continuous from above and decreasing one . □

The concept of ordered fuzzy numbers (OFN) was introduced by Kosiński and his co-writers [9] as an extension of the concept of LR-type FN. Thus, any OFN should be determined by (1) as a fuzzy subset in the real line $\mathbb{R}$. On the other hand, Kosiński has defined OFN as a ordered pair of functions from the unit interval $[0,1]$ into $\mathbb{R}$. This kind of pair is not a fuzzy subset in $\mathbb{R}$. Thus we can not accept original Kosiński's terminology. What is more, the intuitive Kosiński's approach to the notion of OFN is very useful. For these

[1] Poznań University of Economics and Business, Department of Investment and Real Estate, al. Niepodległości 10, 61-875 Poznań, Poland, krzysztof.piasecki@ue.poznan.pl.

reasons, below we present a revised definition of OFN which fully corresponds to the intuitive definition by Kosiński. The OFN concept of a number is closely linked to the following ordered pair.

Definition 2. By the Kosiński's pair $[\![S]\!]$ we understand the ordered pair (f_S, g_S) of continuous weakly monotonous surjections $f_S: [0,1] \to UP_S$ and $g_S: [0,1] \to DOWN_S$ fulfilling the conditions:

$$\big(f_S(1) - f_S(0)\big) \cdot \big(g_S(1) - g_S(0)\big) \leq 0\,, \tag{2}$$

$$|f_S(1) - g_S(1)| \leq |f_S(0) - g_S(0)|, \tag{3}$$

$$UP_S \cap DOWN_S = \{f_S(1)\} \cap \{g_S(1)\}. \quad \square \tag{4}$$

For any Kosiński's pair (f_S, g_S) the function $f_S: [0,1] \to UP_S$ is called the up-function. Then the function $g_S: [0,1] \to DOWN_S$ is called down-function. The up-function and down-function are collectively referred as Kosiński's maps. For any nondecreasing Kosiński's map l_S we define its generalized inverse function as follows

$$\forall_{x \in [l_S(0), l_S(1)]}: \quad l_S^{\triangleleft}(x) = \max\{\alpha \in [0;1]: l_S(\alpha) = x\}, \tag{5}$$

In analogous way, for any nonincreasing Kosiński's map r_S we define its generalized inverse function as follows

$$\forall_{x \in [r_S(1), r_S(0)]}: \quad r_S^{\triangleright}(x) = \min\{\alpha \in [0;1]: r_S(\alpha) = x\}\,. \tag{6}$$

The condition (2) implies that Kosiński's maps cannot be increasing or decreasing at the same time. Knowing this fact, we define OFN in following way.

Definition 3. For fixed Kosiński's pair $[\![S]\!]$ the OFN $\overleftrightarrow{S}$ is defined as the pair of g-LR-type FN and orientation in that the following way:

- left reference function L_S is equal to generalized inverse of the nondecreasing Kosiński's map;
- right reference function R_S is equal to generalized inverse of the nonincreasing Kosiński's map;
- orientation is determined as common sense of all vectors from up-function range UP_S to down-function range $DOWN_S$. $\square$

The above definition is coherent to the intuitive Kosiński's approach to the notion of OFN. Therefore any OFN may be called the Kosiński's number. The space of all OFN is denoted by the symbol $\mathfrak{K}$.

The continuity of Kosiński maps implies that UP_S and $DOWN_S$ are bounded intervals. Numbers $f_S(0)$ and $f_S(1)$ are boundaries of the interval UP_S. Numbers $g_S(0)$ and $g_S(1)$ are boundaries of the interval $DOWN_S$. Therefore, we will denote any OFN $\overleftrightarrow{S}$ with given UP_S and $DOWN_S$ by the symbol $\overleftrightarrow{S}\big(f_S(0), f_S(1), g_S(1), g_S(0)\big)$. The conditions (2), (3) and (4) imply that this OFN fulfils exactly one from the following conditions:

$$f_S(0) \leq f_S(1) \leq g_S(1) \leq g_S(0), \tag{7}$$

$$f_S(0) \geq f_S(1) \geq g_S(1) \geq g_S(0), \tag{8}$$

If $f_S(0) < g_S(0)$, then the condition (7) describes the positive orientation of OFN. In this case, the up-function f_S is nondecreasing and the down-function g_S is nonincreasing. Moreover, positive oriented OFN $\overleftrightarrow{S}\big(f_S(0), f_S(1), g_S(1), g_S(0)\big)$ explicitly determines FN $S\big(f_S(0), f_S(1), g_S(1), g_S(0)\big)$ described by its membership function $\mu_S \in [0;1]^{\mathbb{R}}$ given as follows

$$\mu_S(x) = \begin{cases} 0 & x < f_S(0) \\ f_S^{\triangleleft}(x) & f_S(0) \leq x \leq f_S(1) \\ 1 & f_S(1) \leq x \leq g_S(1), \\ g_S^{\triangleright}(x) & g_S(1) \leq x \leq g_S(0) \\ 0 & g_S(0) < x \end{cases} \tag{9}$$

If $f_S(0) > g_S(0)$, then the condition (8) describes negative orientation of OFN. In this case, the up-function f_S is nonincreasing and the down-function g_S is nondecreasing. Negative oriented OFN $\overleftrightarrow{S}\big(f_S(0), f_S(1), g_S(1), g_S(0)\big)$ explicitly determines FN $S\big(g_S(0), g_S(1), f_S(1), f_S(0)\big)$ described by its membership function $\mu_S \in [0;1]^{\mathbb{R}}$ given as follows

$$\mu_S(x) = \begin{cases} 0 & x < g_S(0) \\ g_S^{\triangleleft}(x) & g_S(0) \leq x \leq g_S(1) \\ 1 & g_S(1) \leq x \leq f_S(1), \\ f_S^{\triangleright}(x) & f_S(1) \leq x \leq f_S(0) \\ 0 & f_S(0) < x \end{cases} \tag{10}$$

For the case $f_S(0) = g_S(0) = a$, the OFN orientation is undefined. We can consider $\overleftrightarrow{S}(a,a,a,a)$ as both positively oriented and negatively oriented. Moreover, $\overleftrightarrow{S}(a,a,a,a)$ represents the crisp number $a \in \mathbb{R}$.

The positivity of OFN is defined in following way

$$\overleftrightarrow{S}\big(f_S(0), f_S(1), g_S(1), g_S(0)\big) > 0 \Leftrightarrow \min\{f_S(0), g_S(0)\} > 0\,. \tag{11}$$

Arithmetic operations on OFN are defined by Kosiński [9] as extension of arithmetic operations on the real numbers. In a special case, for any monotonic function $h: \mathbb{R} \supset \mathbb{A} \to \mathbb{R}$ and OFN $\overrightarrow{S} \in \mathfrak{K}$ we have

$$\overrightarrow{Z} = h(\overrightarrow{S}) \tag{12}$$

where $\overrightarrow{Z} \in \mathfrak{K}$ and

$$\forall_{x\in[0,1]}: f_Z(x) = h(f_S(x)) \wedge g_Z(x) = h(g_A(x)). \tag{13}$$

3 Ordered fuzzy present value

Let us consider the fixed security. We observe a market price $\check{C} > 0$ of this security. In agree with asumption given in [6], PV is such a positive fuzzy number, which is an approximation of the market price $\check{C}$. Therefore, we can determine PV as positive LR-type FN dodatnią $Pv(\check{C}_{min}, \check{C}_*, \check{C}^*, \check{C}_{max})$, where

- $0 < \check{C}_{\min}$ is the maximal lower evaluation of the possible market price,
- $\check{C}_*$ is the minimal upper evaluation of such prices which are noticeably lower than the observed market price $\check{C}$,
- $\check{C}^*$ is the maximal lover evaluation of such prices which are noticeably lower than the observed market price $\check{C}$,
- $\check{C}_{\max}$ is the minimal upper evaluation of the possible market price.

It is obvious that we additionally have

$$\check{C}_{min} \le \check{C}_* \le \check{C}, \tag{14}$$

$$\check{C} \le \check{C}^* \le \check{C}_{max} \tag{15}$$

For given left reference function $L_{Pv} \in [0,1]^{[\check{C}_{min}, \check{C}_*]}$ and right reference function $R_{Pv} \in [0,1]^{[\check{C}^*, \check{C}_{max}]}$, the membership function $\mu_{Pv} \in [0,1]^{\mathbb{R}}$ of LR-type FN $Pv(\check{C}_{min}, \check{C}_*, \check{C}^*, \check{C}_{max})$ is uniquely defined by identity (1). In this section the information described by defined above PV is supplemented with a subjective forecast of orientation of the market price trend. This forecast was implemented in the model BPV as an orientation of fuzzy number. In agree with [10], we use here the following rules recording the alleged orientation of the trend:

1. The presumption of market price increase is described by the positive orientation of fuzzy number. Then PV will be presented as OFN $\overrightarrow{Pv}(\check{C}_{min}, \check{C}_*, \check{C}^*, \check{C}_{max})$
2. The presumption of market price decrease is described by the negative orientation of fuzzy number. Then PV will be presented as OFN $\overleftarrow{Pv}(\check{C}_{max}, \check{C}_*, \check{C}^*, \check{C}_{min})$.

In this way PV is presented as the ordered fuzzy number. Each of these PV representations is called oriented fuzzy PV (OFPV). By writing OFPV with any orientation we will denote it by the symbol $\overleftrightarrow{Pv}(f_{Pv}(0), f_{Pv}(1), g_{Pv}(1), g_{Pv}(0))$, where ordered pair (f_{Pv}, g_{Pv}) is a Kosiński's pair. Such defined OFPV is applied for determining the return rate.

4 Ordered fuzzy return rate

Let us assume that the time horizon $t > 0$ of an investment is fixed. Then, the security considered here is determined by two values:

- anticipated FV $V_t \in \mathbb{R}^+$,
- assessed PV $V_0 \in \mathbb{R}^+$.

The basic characteristic of benefits from owning this security is a return rate r_t given by the identity

$$r_t = r(V_0, V_t). \tag{16}$$

In the general case, if $(V_0, V_t) \in \mathbb{R}^+ \times \mathbb{R}^+$ then the function: $r: \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ is a decreasing function of PV and an increasing function of FV. It implies that for any pair $(V_0, V_t) \in \mathbb{R}^+ \times \mathbb{R}^+$ we can determine inverse function $r_V^{-1}(V_0, \cdot): \mathbb{R} \to \mathbb{R}^+$. Moreover, in the special case we have here:

- simple return rate

$$r_t = \frac{V_t - V_0}{V_0} = \frac{V_t}{V_0} - 1 \tag{17}$$

- logarithmic return rate

$$r_t = \ln \frac{V_t}{V_0}. \tag{18}$$

By the term of a security we understand the authorization to receive future financial revenue, payable to a certain maturity. The value of this revenue is interpreted as anticipated FV of the security. According to the uncertainty theory introduced by Mises [13] and Kaplan [5], anyone unknown to us the future state of affairs is uncertain. The Mises-Kaplan uncertainty is a result of our lack of knowledge about the future state of affairs. Yet, in the researched case, we can point out this particular time in the future, in which the considered state of affairs will be already known to the observer. This kind of Mises-Kaplan uncertainty will be further referred to as "uncertainty". Behind Kolmogorov [7, 8], Mises [12], Lambalgen [10], Sadowski [16], Czerwiński [2], and Caplan [1] we will accept that this is a sufficient condition for modelling the uncertainty with probability. Thus, the uncertainty is often also called a quantitative uncertainty. It is worth noting that FV is not burdened by the Knight's uncertainty [6]. All this leads to the conclusion that FV is a random variable random variable $\tilde{V}_t: \Omega \longrightarrow \mathbb{R}^+$. The set Ω is a set of elementary states ω of the financial market. In the classical approach to the problem of the return rate estimation, the security PV is identified with the observed market price $\check{C}$.Then the return rate is a random variable determined by the identity

$$\tilde{r}_t(\omega) = r\left(\check{C}, \tilde{V}_t(\omega)\right). \tag{19}$$

In practice of financial markets analysis, the uncertainty risk is usually described by probability distribution of return rate determined by (19). At the moment, we have an extensive knowledge on this subject. Let us assume that this probability distribution is given by cumulative distribution function $F_r: \mathbb{R} \longrightarrow [0;1]$. We assume here that the expected value $\bar{r}$ of this distribution exists. On other side, the cumulative distribution function F_r determines probability distribution $P: 2^{\Omega} \supset \tilde{r}^{-1}(\mathcal{B}) \longrightarrow [0;1]$, where the symbol $\mathcal{B}$ denotes the smallest Borel σ-field containing all intervals in the real line $\mathbb{R}$. Moreover, let us note that there we have

$$\tilde{V}_t(\omega) = r_V^{-1}\left(\check{C}, \tilde{r}_t(\omega)\right) \tag{20}$$

Let us consider now the case when PV is determined as OFPV $\overleftarrow{Pv}$ represented by the Kosiński's pair (f_{Pv}, g_{Pv}). Then using (16) we obtain ordered fuzzy return rate (OFRR). In agree with (13) and (16), for any $\omega \in \Omega$ of financial market, we assign the OFRR $\overrightarrow{\mathcal{R}}(\omega)$ to considered security. Assigned OFRR $\overrightarrow{\mathcal{R}}(\omega)$is represented by the Kosiński pair $(f_{\mathcal{R}}(\cdot\,|\omega), g_{\mathcal{R}}(\cdot\,|\omega))$. Due (13), (16) and (20), the Kosiński's functions listed here are determined by the identities

$$f_{\mathcal{R}}(x|\omega) = r\left(f_{Pv}(x), \tilde{V}_t(\omega)\right) = r\left(f_{Pv}(x), r_V^{-1}\left(\check{C}, \tilde{r}_t(\omega)\right)\right), \tag{21}$$

$$g_{\mathcal{R}}(x|\omega) = r\left(g_{Pv}(x), \tilde{V}_t(\omega)\right) = r\left(g_{Pv}(x), r_V^{-1}\left(\check{C}, \tilde{r}_t(\omega)\right)\right). \tag{22}$$

For any $y \in \mathbb{R}^+$ the function $r\left(y, r_V^{-1}(\check{C}, \cdot)\right): \mathbb{R} \to \mathbb{R}^+$ is decreasing. It implies that this function is measurable. Therefore the expected OFRR $\overrightarrow{R}$ is OFN described by means of the Kosiński' pair (f_R, g_R) determined in the following way

$$f_R(x) = \int_{\Omega} f_{\mathcal{R}}(\cdot\,|\omega)\mathrm{d}P = \int_{-\infty}^{+\infty} r\left(f_{Pv}(x), r_V^{-1}(\check{C}, y)\right) dF_r(y) = r\left(f_{Pv}(x), r_V^{-1}(\check{C}, \bar{r})\right), \tag{23}$$

$$g_R(x) = \int_{\Omega} g_{\mathcal{R}}(\cdot\,|\omega)dP = \int_{-\infty}^{+\infty} r\left(g_{Pv}(x), r_V^{-1}(\check{C}, y)\right) dF_r(y) = r\left(g_{Pv}(x), r_V^{-1}(\check{C}, \bar{r})\right). \tag{24}$$

where the symbol $\bar{r}$ denotes expected value of return rate (19). Since $r: \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ is decreasing function of the first argument we can say that for any $x \in [0;1]$ we have

$$f_R(x) \geq g_R(x) \Leftrightarrow f_{Pv}(x) \leq g_{Pv}(x). \tag{25}$$

It means that OFPV and determined by it expected OFRR always have opposite orientations. Then we can say that

- predicted increase in PV allows us to anticipate a decrease in the expected return rate,
- predicted decrease in PV allows us to anticipate an increase in the expected return rate.

In theory and practice of finance both of these facts are well known. . This observation proves that the extension of the fuzzy model of imprecise estimates of PV and the return rate to description by means of OFN is the appropriate direction for the development of fuzzy finance theory.

Moreover, in special case we have:

– expected OFRR $\overleftrightarrow{R}_t$ represented by the Kosiński's pair (f_R, g_R) determined as follows

$$f_R(x) = \frac{\check{C}\cdot(1+\bar{r})}{f_{Pv}(x)} - 1, \tag{26}$$

$$g_R(x) = \frac{\check{C}\cdot(1+\bar{r})}{g_{Pv}(x)} - 1\ ; \tag{27}$$

– expected logarithmic OFRR $\overleftrightarrow{\hat{R}}_t$ represented by the Kosiński's pair $(f_{\hat{R}}, g_{\hat{R}})$ determined as follows

$$f_{\hat{R}}(x) = \ln\left(\frac{\check{C}\cdot e^{\bar{r}}}{f_{Pv}(x)}\right), \tag{28}$$

$$g_{\hat{R}}(x) = \ln\left(\frac{\check{C}\cdot e^{\bar{r}}}{g_{Pv}(x)}\right). \tag{29}$$

The formal simplicity of the above relationships is another benefit which we gain from the use of OFNs for imprecise estimation of return rates. The simplicity of these relationships is also good for further development of the portfolio theory.

Acknowledgements

The results of the work fully convince that the use of OFN will facilitate the analysis of financial instruments with imprecise estimated values. It is expedient to further development the fuzzy finance theory based on OFN. In case of analysis of a single financial instrument, we can adopt here the methods recommended in [15, 16]. For construction of a portfolio of imprecise evaluated financial instruments, we can also use the models of portfolio construction recommended in [17].

Such broad possibilities encourage further research into the application of ordered fuzzy numbers in the theory and practice of quantified finance.

References

[1] Caplan B.: Probability, common sense, and realism: a reply to Hulsmann and Block, *The Quarterly Journal of Austrian Economics*, **4(2)**, (2001), 69-86
[2] Czerwiński Z.: Enumerative induction and the theory of games, *Studia logica*, **10** (1960), 29-38.
[3] Dubois D., Prade H.: *Fuzzy real algebra: some results*, *Fuzzy Sets and Systems* **2**, (1979), 327-348.
[4] Dubois D., Prade H.: *Fuzzy set and systems: theory and applications*, Academic Press, New York, 1980.
[5] Kaplan S., Barish N.N.: Decision-Making Allowing Uncertainty of Future Investment Opportunities, *Management Science*, **13(10)**, (1967), 569–577.
[6] Knight, F. H.: Risk, Uncertainty, and Profit, Hart, Schaffner & Marx; Houghton Mifflin Company, Boston, MA, 1921.
[7] Kolmogorov, A. N.: *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Julius Springer, Berlin, 1933
[8] Kolmogorov A. N.: *Foundations of the theory of probability*, Chelsea Publishing Company, New York, 1956.
[9] Kosiński W., Prokopowicz P., Ślęzak D.: Ordered fuzzy numbers. *Bulletin of the Polish Academy of Sciences, Sér. Sci. Math*., **51 (3)**, (2003), 327-338.
[10] van Lambalgen M.: Randomness and foundations of probability: von Mises' axiomatization of random sequences, In: *Probability, statistics and game theory, papers in honor of David Blackwell*,(Ferguson T.S., Shapley L.S., Mac Queen J.B. ,eds.) Institute for Mathematical Statistics, Harward, 1996, 347-368.
[11] Łyczkowska-Hanćkowiak A.: Behavioural Present Value determined by ordered fuzzy number, *SSRN Electronic Journal* (2017), DOI:10.2139/ssrn.2988243
[12] Mises R. von: *Probability, statistics and truth*, The Macmillan Company, New York, 1957.
[13] Mises L. von: *The Ultimate Foundation of Economic Science An Essay on Method*, D. Van Nostrand Company, Inc., Princeton, 1962.
[14] Piasecki K.: Behavioural Present Value, *SSRN Electronic Journal* (2011), DOI:10.2139/ssrn.1729351
[15] Piasecki, K.: Effectiveness of securities with fuzzy probabilistic return, *Operations Research and Decisions* **21(2)**, (2011), 65-78.
[16] Piasecki K.: On Imprecise Investment Recommendations, *Studies in Logic, Grammar and Rhetoric*, **37(50)** (2014), 179-194
[17] Piasecki K., Siwek J.: Two-asset portfolio with triangular fuzzy present values - an alternative approach, submitted to Proceedings of *International Conference On Accounting, Finance And Financial Institutions. Theory Meets Practice*, 2016
[18] Sadowski W.: Forecasting and decision making, In: *Quantitative Wirtschafts- und Unternehmensforschung*, (Henn R., Schips B., Stähly P., eds.)Springer-Verlag, Berlin Heidelberg, 1980, 92-102

Use of multi-criteria decision analysis in fuzzy network DEA models

Michal Pieter[1]

Abstract. Data Envelopment Analysis (DEA) is a popular tool for examining the efficiency of production units in transforming a set of inputs into outputs; however, traditional models consider them to be closed, black-box systems. Network DEA (NDEA) allows one to capture the inner structure as a series of interacting processes, making for a more precise model of complex systems.
In an effort to further increase the accuracy of real world representation in models, several directions of research into NDEA have been developed. One of them replaces crisp data with fuzzy values to better capture the imprecise nature of most complex systems. Elsewhere, models that use a number of Multi-criteria Decision Analysis (MCDA) methods allow managers to put restrictions on allowable input/output weights, to rank efficient units, or to solve other scenarios for which traditional models are ill-equipped.
This paper aims to first provide an overview of existing models utilizing both aforementioned approaches. Subsequently, further ways in which both approaches could be combined to yield brand new methods are examined. Advantages and limitations of such endeavor, as well as proposed practical applications of this combined technique, follow.

Keywords: DEA, NDEA, network, data envelopment analysis, fuzzy, multi-criteria decision analysis, MCDA.

JEL Classification: C61, C67
AMS Classification: 90-02, 90-06, 90B50, 90C08, 90C29, 90C70

1 Introduction

Ever since its inception in the late seventies, data envelopment analysis (DEA) has been a useful and popular tool for comparing efficiency of units in a homogenous group. These units (decision-making unit, DMU) produce certain outputs by consuming and transforming various inputs and the way in which the two sets relate determine their efficiency. The first model in the family was put forward by Charnes, Cooper and Rhodes [4] mainly as a way to determine the efficiency of government agencies and non-profit organizations in absence of clear-cut metrics such as profit. Since then, the scope of application has expanded greatly in both numbers and diversity, ranging from power grids to environmental project assessment, and of course private companies. Over the years, improvements and expansions have been made to the original model. Banker, Charnes and Cooper [1] allowed for variable returns to scale. Tone [19] then introduced a new approach, optimizing both inputs and outputs at the same time. These models allow for identification of efficient units and to rank those not efficient. Therefore, various models have been constructed that allow full ranking of units, including so-called super-efficiency models, such as [20] with their over-unity efficiency scores. All of these, however, treat DMUs as indivisible and transparent entities with inputs coming in, outputs coming out, but not being able to say anything about the transformation itself. In most real world scenarios, that is a far too simplistic approach. Thus, the need to see inside the DMU arose.

1.1 Network data envelopment analysis

Some of the first forays into what would later become known as network DEA (NDEA) modeled the internal structure of units as consisting of two *processes* (sub-processes, subunits), where outputs of the first process would serve as inputs of the second. This is known as a series structure. Färe and Grosskopf [5], and later others, expanded the notion of multiple processes to a parallel structure, where processes don't depend on each other but share the same inputs, as well as a mixed structure of both. This structure, as well as additional requirements, such as whether intermediate outputs of a given process can serve as final output of the unit, determine the precise form of the mathematical model. All processes are generally assumed to be characterized by the same inputs and outputs – if that is not the case, the vectors are transformed and irrelevant inputs and outputs are required to be zero.

[1] University of Economics Prague, Department of econometrics, W. Churchill sq. 4, 130 67 Praha 3, Czech Republic, xpiem02@vse.cz.

In traditional models, the performance of a unit in turning inputs into outputs – it's efficiency – is measured using the unit's distance in vector space to the production frontier, which is in turn defined by the set of efficient units. How the production frontier is constructed depends on the model, notably by the nature of returns to scale. The precise way in which the distance is measured also depends on the model – originally, input- and output-oriented versions of the model had to be used separately. Later, Tone [19] proposed a model that combines both. In case of multiple processes, the situation is more complicated – not only the way to measure efficiency of the processes needs to be decided, but how to determine the overall efficiency of the unit as well. Several broad approaches [6] to this problem have been proposed, e.g. by weighing processes to obtain the system average.

1.2 Fuzzy data in network DEA

Generally, when formulating linear programming problems, there is an unspoken assumption that all the available data is either *known with certainty*, or any implicit uncertainty has been replaced with educated guesswork and is not further considered. Thus, the model operates throughout with precise, so called crisp values, from the initial data all the way to computed results. There are cases where data can be gathered with such a high degree of confidence, that this assumption can be considered to hold. However, in real world scenarios, ordinal values, categories and confidence intervals are much more common. It would clearly be advantageous to include this *imprecise nature* of the real system into the model as well. This data is described as being fuzzy and modification of existing models to accommodate it is therefore known as fuzzification.

In the case of fuzzy models, all inputs and outputs are considered to be either fuzzy numbers, or more generally fuzzy intervals, defined by a membership function. Crisp values can be considered a degenerate form thereof, so they can be freely mixed in as well. Kao and Liu [8] present a model where two processes are connected in series and where data is entered in the form of triangular fuzzy numbers. This leads to a non-linear programming problem, whose solution gives the efficiencies of both processes and the unit as a whole, again in the form of a triangular fuzzy number. Moreover, the efficiency of the unit is a *product of the process efficiencies*. In [7], a similar approach is used to examine units with parallel structure. Shermeh et al. [17] use a modified radial network model in which a parameter can be set to determine the "fuzziness" of the model. As in the previous cases, the efficiency is returned in the form of an interval fuzzy number, but in this case is given as a *weighted average* of process efficiencies.

1.3 Multi-criteria decision analysis in traditional models

When considering possible solutions to a given problem, it is often useful to consider multidisciplinary approaches. The case with DEA is no different and many researchers have turned to related fields in search of further extensions of traditional models. One of these is Multi-criteria decision analysis (MCDA, also decision-making, MCDM), a branch of decision theory providing many methods for selecting one option out of many, ranking of the options, all based on a set of criteria. In this respect, it shares some similarity with DEA. For example, a two-phase approach combining CCR DEA model with AHP method allows for full ranking of units, as shown in [18]. A major branch of MCDA – multi-objective programming – also enables one to optimize several objective functions at once.

There are several options in combining both fields of study. DEA can be used in a supportive role to some MCDA method. Common problem in many methods is how to properly weigh the different criteria. Several approaches, including the aforementioned, use a DEA model to replace subjective valuation with the computed optimal weights of criteria playing the role of inputs and outputs. On the other hand, sometimes it is reasonable in a DEA model to require that there are restrictions on how an input or output of certain unit can be weighed in the optimal solution. Various MCDA methods are then well suited to elicit these restrictions based on decision-maker's preferences, as is the case in [21]. Other examples, as well as a thorough comparison can be found in [2].

2 Overview of existing approaches

Khalili-Damghani and Shahmir [10] offer one example on how to combine both fields, as outlined in the previous chapter. In order to study the efficiency of Iranian power production plants and distribution networks, they first consider a network DEA structure, where each DMU is composed of a series-connected pair of processes, representing production and distribution of electric power, respectively. Undesirable outputs of the first subunit, as well as *exogenous inputs* of the second, in addition to the intermediate ones outputted by the first subunit, were also built into the model. This leads to three non-linear programming problems, so far with crisp data, one to compute the crisp efficiency of each subunit and one of the DMU as a whole. Linearization of models is then achieved by employing the well-known Charnes-Cooper transformation. The element of uncertainty is then implemented by considering *interval numbers* – concept closely linked with fuzzy set theory – instead of crisp values. For each observed value of input and output a lower and upper bound is provided. Likewise, the calculated efficiencies are also given as an interval, where lower and upper bounds represent the *pessimistic* and *optimistic* scenario, respectively. In an optimistic scenario, the studied DMU is assumed to achieve the best possible performance, while all

the others are at their worst, and in the pessimistic scenario, the reverse is true. Based on these bounds, each DMU or its subunit can be put into one of three categories. Those with both bounds at unity are always efficient, those with both less than one are always inefficient, and all others can be either efficient or not. The key use of MCDA methods lies in the utilization of TOPSIS method to rank the DMUs and subunits. Calculated lower and upper bounds are used to construct basal and ideal alternative, to which they are then compared, yielding two distance values – d^- and d^+, respectively. Finally, closeness coefficient, given as d^-/d^-+d^+, is computed, allowing the DMUs and both subunits to be ranked, based on its decreasing value. Thus, this approach is reminiscent of earlier use of MCDA methods to offer ranking of units where DEA models by themselves are unable to do so effectively.

Wang et al. [22] utilize a unique approach in their model to study holdings in the banking sector. Theirs is a two-stage network DEA model with a series structure, this time however with no undesirable first-stage outputs or exogenous second-stage inputs. All outputs from the first process – profitability – are immediately "consumed" by the second – creation of value. A multi-objective model is constructed, where each objective function represents a DMU trying to maximize its overall efficiency, using a common set of weights for inputs, outputs and intermediaries. Uniquely, only then is fuzziness introduced in an effort to *transform* multi-objective model into a single-objective one. The multi-objective efficiency of any given DMU j is taken to be a fuzzy number, whose membership function $f_j(\theta_j)$ is defined with respect to maximum and minimum efficiency values of that DMU, were they optimized on their own. Once the objective functions are replaced with these fuzzy efficiency values, they propose to transform the model by maximizing the least efficient DMU – the maximin approach. Since both the efficiency θ_j and the membership function attain values from 0 to 1, model can be simplified with $f_j(\theta_j)= \theta_j$. By solving the resulting non-linear model, optimal common weights can be used to compute overall efficiency for each DMU, which can be further decomposed to a *product* of the first-stage and second-stage efficiency values. The authors claim that their fuzzy multi-objective approach is simpler to calculate – presumably because only one model is solved for the entire set of DMUs – and allows for better discrimination of DMUs and more accurate identification of the sources of inefficiencies, than traditional models.

Kao et al. [9] offer a similar approach to that of Wang et al. They consider a general case of k sub-processes, connected in series, with possible exogenous inputs. In the multi-objective model, each sub-process tries to maximize its particular efficiency in an objective function, while the same is true for the DMU, whose efficiency is defined as a *weighted sum* of sub-process efficiencies, as opposed to their product. The resulting model has thus k+1 objective functions and studies just the one DMU, as opposed to all of them. Otherwise, the computational steps are shared between the models, including the way the membership function of each sub-process is formed, using an ideal and basal efficiency scores. Likewise, minimax is used to transform the model into a single-objective one, even though a possible alternative approach is presented – namely, maximization of weighted averages. In any case, by solving the model, the efficiency, as well as weight of each sub-process is obtained and subsequently aggregated into DMU efficiency. It is clear that no reduction in computational resources has been reached and is more suited to DMU intra-analysis, rather than comparison. However, it allows one to consider an *arbitrary* number of subunits and calculate their efficiencies, as well as how they contribute to the efficiency of the DMU itself.

3 Possibilities for further research

As evidenced by examples in chapter 2 the existing research in this particular area of study is limited at best. This can be considered surprising, mainly because considerable effort is being made in combining two of the necessary component approaches. In fact, especially in recent years a multitude of papers have been published on fuzzy network DEA and on use of various MCDA methods (such as TOPSIS, PROMETHEE, etc.) in either fuzzy DEA *or* network DEA. It may be useful to briefly consider why that may be so. One possibility is that full combination of all the approaches offers few advantages in practical situation. This position seems untenable – each of the extensions solves a particular problem of traditional models, but does nothing to address the other ones. Network DEA peers inside the DMU, fuzzification captures the imprecise nature of real systems and utilization of MCDA can help for example with weight restriction and/or ranking of efficient units. This, together with the usefulness of piecewise combination of these three approaches, as evidenced by the abundance of literature, proves that a full combination must offer some kind of advantage in the real world, if only it were available. Another possibility is a lack of academic interest. This also seems unlikely, given the aforementioned and also existing, albeit few, papers on the subject. It would seem that a natural progression is at play. Many of the piecewise combinations have been put forward in just recent years, so it's only reasonable that the more complex ones should only now start to appear.

To examine possible new directions of research, it's useful to differentiate between two major branches of multi-criteria decision analysis. MCDA acompasses two similar, yet in one fundamental way very different fields. Multi-criteria evaluation deals with discrete options while multi-objective programming (MOP) with continuous functions. In chapter 2, we saw both of them utilized to solve a particular issue in fuzzy network DEA models. What follows are some of the ways they can be used to yield new combined models.

3.1 Utilization of multi-criteria evaluation

Let us first consider the ways in which multi-criteria evaluation can be utilized. An obvious approach is to examine its use in more basic models and find out if and how it can be applied to DEA models that are both fuzzy and with network structure. One such case is *weight restriction*. One of the strengths of traditional DEA is that each DMU can choose weights of inputs or outputs so that its efficiency is highest. This can have side effects, though. Too many DMUs can be identified as efficient, which prevents discriminating them from one another – however, some models, such as super-efficiency [20], solve this. In other cases, the chosen weights can be impossible in reality or unacceptable in practice. Thus, restricting their allowable range may be necessary. Vieira Junior [21] proposed using a MACBETH method to restrict allowable weights for the observed DMU. A decision-maker would compare each pair of inputs (the same with outputs) on a linguistic scale. This information would then be turned into lower and upper bounds for each input and output weight. The proposed model ensures that they are within the allowed range. Other approaches can be used to restrict the weights as well, e.g. the ANP based approach in [11]. In network DEA, weight restriction may be even more useful. In parallel structure, it may be necessary to regulate the distribution of inputs between competing subunits. In a series structure, weight restriction of intermediate factors may be necessary in the presence of either exogenous inputs to sub-processes or of double-role intermediate products (that can also serve as final outputs). One could apply the same methods to obtain the allowable ranges of these weights as well and then modify the model accordingly. One existing example is Farzipour Saen [16] introducing a two-stage model with weight-restricted dual-role factors. Because weights are crisp values even in fuzzified versions of the models, they can be bounded in these as well. In fact, Liu [13] proposes a fuzzy model with weight restrictions for the simplest case of two-stage series network DEA with no exogenous intermediate inputs – the only difference is the method of obtaining the restrictions. One could conceivably construct a generalized model that utilizes either the aforementioned MACBETH, or some other multi-criteria evaluation method.

While the use of fuzzy data improves the discriminatory power of the models, there may still arise situations where several DMUs or their subunits are identified as "almost equally" efficient, to keep in line with fuzzy terminology. In case of triangular fuzzy numbers, let us consider the case of one DMU having an efficiency (also a fuzzy number) whose membership function overlaps or is nested within that of another DMU. One could use increasing values of α-cuts to discriminate between them, but that may give a distorted impression, e.g. because of greater spread in one of the functions. In chapter 2 we saw the use of TOPSIS method to rank DMUs. Sinuany-Stern et al. [18] proposed a two-phase AHP-DEA approach that consists of pairwise comparisons of each DMU in constructing an efficiency matrix. Single-level AHP is then used to compute the weight of each DMU and rank them accordingly. Similar approach could be applied to the offending DMUs or subunits in a fuzzy environment. If the number of DMUs is not large, one might also consider pairwise comparisons of the entire set. Of course, a fuzzy version of AHP method would need to be utilized, whose methodology can be found for example in [3].

3.2 Utilization of multi-objective programming

In chapter 2, we saw two papers that utilize multi-objective programming in a fuzzy network DEA context. There, each objective function represented the efficiency of either DMU or its subunits. Multi-criteria DEA (MCDEA), as proposed by Li and Reeves [12], offers a similar approach, however, here the objective functions represent different conceptions of efficiency itself. Namely, it considers deviation from the efficient solution (inefficiency) and then tries to minimize this deviation for the observed DMU; the maximum deviation of all DMUs; and the aggregate deviation of all DMUs. By using what is essentially a *lexicographic* approach, one can improve the discrimination power over traditional DEA, represented by the first objective. Mahdiloo et al. [14] combined this very approach with the simplest network model (two-stage series structure, no intermediate exogenous inputs or double-role factors). In the objective functions they consider the mean of the sub-process deviations and then, instead of lexicographic approach, transform the model into a single-objective by *weighting* the three goals. By carefully applying the same principles to the fuzzy two-stage model proposed by Kao and Liu [8] one could arrive at an as of yet unexplored fuzzy two-stage MCDEA model. However, this would invariably result in a pair of non-linear models of considerable complexity. Given that fuzzification itself helps with DMU discrimination, the costs-to-benefit ratio of this approach may or may not be favorable and would need to be examined separately.

4 Practical considerations

Fuzzy network DEA models, even in their most basic form – two-stage series structure and triangular fuzzy numbers – are in themselves rather complicated and often lead to non-linear programming problems. Introduction of additional complexity through the use of MCDA, e.g. as proposed in chapter 3, is arguably the major drawback of such approach. Computational difficulty of the resulting model must be especially considered when proposing any new models, with respect to the benefits it may provide. In some models, especially the weight restriction ones, utilizing multi-criteria evaluation methods, subjective information provided by a group of experts is necessary to

elicit weight bounds for individual factors. In some circumstances this might be considered a weakness. However, in the case of weight-restriction, this subjective viewpoint is uniquely suited to disallow impossible or impractical factor weights, with the discrimination issue already solved mainly through the use of fuzzy data. In the case of multi-objective programming, its main benefit lies in the *simplification* of the resulting model, allowing one to optimize either various components of DMUs or different conceptions of efficiency, all at once.

It is clear that all three approaches strive for the same ultimate goal – to better approximate the real system in the model. The main benefit of their combination then lies it the fact that they do so through complementary means. While network DEA allows one to consider component parts of the system, normally hidden inside the DMU, fuzzy data captures the imprecise nature present in many complex systems. Various MCDA approaches then allow one to do away with unrealistic values of inputs or outputs in the optimal solution, or to consider multiple competing goals at the same time. Finally, the combination of MCDA and DEA lends itself to be utilized not only as a means of *evaluation* of DMUs, but also as a *selection* tool for decision-makers.

Practical applications of both the existing and proposed approaches stem precisely from the aforementioned unique combination of strengths, and also limitations. It is obvious that the combined approach can only be applied in systems that are either made of, or can be modelled as being made of, component sub-processes. As mentioned in chapter 2, Khalili-Damghani and Shahmir [10] apply their model to rank electric power companies, in both production and distribution stages. In the banking sector, Wang et al. [22] break down the sustainable operations of a bank holding to creation of profit and creation of value. Kao et al. [9] demonstrate their model on two examples from the electric power industry. As evidenced, electric power and other utilities are well suited for examination by the combined approach. While some factors are easily measured, others, such as environmental impact or quality of service require the use of fuzzy data. Furthermore, presence of large number of factors lends itself well to the use of various MCDA techniques. As one of their examples, Kao et al. [9] measure the efficiency of Taiwanese solar energy companies, consisting of R&D, production and sales subunits. This multiple-stage structure could be easily applied to many other private companies. Similar structure could be also applied to e.g. educational institutions (outreach, education, research) or other public organizations. Beyond these, many other areas could benefit from the proposed models. In an increasingly environmentally conscious world, many papers in both fuzzy and network DEA have been published in recent years on the subject of green supply chain management, i.e. where environmental concerns play a significant role. For example, Mirhedayatian et al. [15] model this chain using 4-stage series-connected (supplier, producer, distributor, consumer) network DEA with fuzzy data. They consider 25 factors, among them CO_2 emissions. In their model, there is a chance that the DMU will place an unacceptably low weight, or none at all, to this factor – one among so many. It would be useful to use an MCDA approach to restrict weights of this undesirable output – as well as others – to an appropriate range and thus ensure that GSCM stay true to its name in prioritizing environmental issues. In conclusion, all of the proposed examples struggle with the same problems – combination of crisp and fuzzy data, large number of various inputs and outputs and multiple sub-processes. In this way, all could benefit from an integrated MCDA and fuzzy network DEA approach.

5 Concluding remarks

Concept of using various MCDA approaches in the particular field of fuzzy network DEA is still rather new and underutilized, as evidenced by brevity of chapter 2. As discussed earlier, this most probably stems from the fact that only relatively recently have researchers begun to combine any two of the approaches. As time goes on, a natural progression towards more complex and integrated solutions will lead to increased interest in the full combination of all three approaches. Even though a small selection already exists, this paper may provide basis for further research in this area. That is, its main contribution lies not so much in the review of existing literature on the subject, sparse as it is, but in examining possible further ways in which new models could be constructed. As we saw in chapter 3, these changes are evolutionary in character, rather than revolutionary. It seems likely that at least in the immediate future, the focus will be on integrating a third part of the MCDA fuzzy network DEA equation with an already existing combination of the other two. Certainly, more papers on the subject will follow.

Acknowledgements

The author would like to extend his gratitude to the Internal Grant Agency of the Faculty of Informatics and Statistics, University of Economics, Prague, that supported this paper through grant no. IGA F4/57/2017.

References

[1] Banker, R.D., Charnes, A., and Cooper, W.W.: Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* **30** (1984), 1078–1092.

[2] Belton, V., and Stewart, T.J.: DEA and MCDA: Competing or Complementary Approaches? In: *Advances in Decision Analysis*, (Meskens, N., and Roubens, M., eds.), Springer Netherlands, 1999, 87–104.

[3] Chang, D.-Y.: Applications of the extent analysis method on fuzzy AHP. *European Journal of Operational Research* **95** (1996), 649–655.
[4] Charnes, A., Cooper, W.W., and Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.
[5] Färe, R., and Grosskopf, S.: Network DEA. *Socio-Economic Planning Sciences* **34** (2000), 35–49.
[6] Kao, C.: Network data envelopment analysis: A review. *European Journal of Operational Research* **239** (2014), 1–16.
[7] Kao, C., and Lin, P.-H.: Efficiency of parallel production systems with fuzzy data. *Fuzzy Sets and Systems* **198** (2012), 83–98.
[8] Kao, C., and Liu, S.-T.: Efficiencies of two-stage systems with fuzzy data. *Fuzzy Sets and Systems* **176** (2011), 20–35.
[9] Kao, H.-Y., Chan, C.-Y., and Wu, D.-J.: A multi-objective programming method for solving network DEA. *Applied Soft Computing* **24** (2014), 406–413.
[10] Khalili-Damghani, K., and Shahmir, Z.: Uncertain Network Data Envelopment Analysis with Undesirable Outputs to Evaluate the Efficiency of Electricity Power Production and Distribution Processes. *Computers & Industrial Engineering* **88** (2015), 131–150.
[11] Kuo, R.J., and Lin, Y.J.: Supplier selection using analytic network process and data envelopment analysis. *International Journal of Production Research* **50** (2012), 2852–2863.
[12] Li, X.-B., and Reeves, G.R.: A multiple criteria approach to data envelopment analysis. *European Journal of Operational Research* **115** (1999), 507–517.
[13] Liu, S.-T.: Restricting Weight Flexibility in Fuzzy Two-stage DEA. *Computers & Industrial Engineering* **74** (2014), 149–160.
[14] Mahdiloo, M., Jafarzadeh, A.H., Saen, R.F., Tatham, P., and Fisher, R.: A multiple criteria approach to two-stage data envelopment analysis. *Transportation Research Part D: Transport and Environment* **46** (2016).
[15] Mirhedayatian, S.M., Azadi, M., and Farzipoor, S.: A novel network data envelopment analysis model for evaluating green supply chain management. *International Journal of Production Economics* **147** (2014), 544–554.
[16] Saen, R.F.: Restricting weights in supplier selection decisions in the presence of dual-role factors. *Applied Mathematical Modelling* **34** (2010), 2820–2830.
[17] Shermeh, H.E., Najafi, S.E., and Alavidoost, M.H.: A novel fuzzy network SBM model for data envelopment analysis: A case study in Iran regional power companies. *Energy* **112** (2016), 686–697.
[18] Sinuany-Stern, Z., Mehrez, A., and Hadad, Y.: An AHP/DEA methodology for ranking decision making units. *International Transactions in Operational Research* **7** (2000), 109–124.
[19] Tone, K.: A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* **130** (2001), 498–509.
[20] Tone, K.: A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research* **143** (2002), 32–41.
[21] Vieira Junior, H.: Multicriteria approach to data envelopment analysis. *Pesquisa Operacional* **28** (2008), 231–242.
[22] Wang, W.-K., Lu, W.-M., and Liu, P.-Y.: A Fuzzy Multi-objective Two-stage DEA Model for Evaluating the Performance of US Bank Holding Companies. *Expert Systems with Applications* **41** (2014), 4290–4297.

Identification of the Direction of Changes in the Structure of Interdependence among the US Capital Market and the Leading European Markets

Michał Bernard Pietrzak[1], Edyta Łaszkiewicz[2], Adam P. Balcerzak[3], Tomáš Meluzín[4], Marek Zincker[5]

Abstract. The aim of the article is to identify global factor (a latent variable) determining direction of changes in the structure of interdependence among US capital market and the leading European markets (UK and Germany). The analysis was done for the years 2004-2014 with daily time series. In the first step, conditional correlations obtained with application of DCC-GARCH model were used to assess the interdependence among the markets. The second stage of the research was devoted to cointegration analysis. For this purpose, stationarity of the conditional correlations was verified with application of Phillips-Perron test. Next, based on the Johansen procedure a long-term relationship structure was established. In the end, a common global factor determining the changes in interdependence was identified with application of confirmatory factor analysis. The common global factor projections reflect the leading direction of changes in the interdependence among the markets. Estimating the quadratic trend model for the identified global factor has confirmed an increase in the interdependence as a result of the global financial crisis and the slow stabilization process of the analyzed markets.

Keywords: capital market, conditional correlations, DCC-GARCH, cointegration, confirmatory factor analysis.

JEL Classification: G12, G15, C58
AMS Classification:91B84

1 Introduction

The last global financial crisis has shown that it is possible for large financial entities involved in the global linkage system to go bankrupt, the best exemplification being the collapse of the Lehman Brothers investment bank. This case proves that with the specific political and institutional conditions, the risk of bankruptcy may also affect financial institutions classified as 'too big to fail'. The collapse of financial markets has confirmed that under globalization it is possible to destabilize national financial systems, including ones in developed economies, which were not originally the place for the development of a speculative bubble that triggered the crisis. This is most evident with regard to Iceland's financial crisis of 2008.

A negative situation occurring on the financial markets is important enough and can move to the real sphere and affect thesocio-economic situation of countries [19]. The influence exerted on the real sphere may become evident in the change of trade relations between countries or within an individual country[30]. The level of foreign direct investment, the level of enterprise investment, and the situation on the labour market are particularly sensitive in that respect [24, 27]. Also, the economic climate for financing innovation potential of the economy may change significantly due to the crisis [33]. The consequences of the existing system of international links vary from country to country, depending on their ability to exploit global connections and the country's resilience to crises. Changes in the international linkages system resulting from crisis situations can have a significant impact on changes in economic policies of countries. As a result of a crisis, countries may radically change their fiscal and monetary policies [5, 17, 20, 23].

For these reasons, the identification of interdependencies in a global linkages system can be considered as one of the key research questions that are needed to identify potential strategies in times of market turbulence

[1]Nicolaus Copernicus University, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: pietrzak@umk.pl.
[2]University of Lodz, ul. Polskiej Organizacji Wojskowej 3/5, 90-255 Łódź, Poland, e-mail: elaszkiewicz@uni.lodz.pl.
[3]Nicolaus Copernicus University, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: adam.balcerzak@umk.pl.
[4] Brno University of Technology, Kolejní 2906/4, 612 00 Brno, Czech Republic, e-mail: meluzint@vutbr.cz.
[5] Brno University of Technology, Kolejní 2906/4, 612 00 Brno, Czech Republic, e-mail: zinecker@fbm.vutbr.cz.

and to develop system tools to improve the functioning of financial systems. This is reflected in the numerous international empirical studies available in the subject literature [1, 2, 6, 7, 10, 15, 16, 22, 29].

This article assumes two research objectives. The first was to characterize the interdependence of selected key markets in the global economic system. Among the most important markets for the global system, the United States capital market is one of the most developed and most important markets in the global financial system. In addition to the US capital market, the study also includes the UK and Germany markets, which are the largest EU capital markets. The second research objective was to identify a global factor common to the researched markets by means of a confirmatory factor analysis. Then, based on the received values of the global factor, a leading direction for the interdependence system between selected capital markets was established.

2 Establishing an interdependence system between the US capital market and major EU markets

According to the assumed objective, we identified interdependencies for the United States, the United Kingdom, and Germany, which allowed us to determine the nature of global interdependencies in these capital markets. The study uses time series for the following stock indices: S&P500 (the United States), FTSE100 (the United Kingdom), and DAX (Germany). The data was downloaded from the website service: http://www.finance.yahoo.com. The study was conducted using daily trading data, taking into accountthe years 2004-2014. The analysis uses the values of the logarithmic return rates calculated for these indices.

The starting point for the analysis was 2004, i.e., the year of the enlargement of the European Union by the new countries. This factor can be regarded as the most important institutional change in the region, which could affect especially German financial market. It should be stressed that the study covered the period of relative economic stability and international pre-crisis prosperity (2004-2007), the period of international instability (2008-2012), and the time of the return to the relative equilibrium of the global financial system (2012-2014). Thus, it can be assumed that the time horizon of the study includes both the period of calm and the period of turbulence caused by the crisis.

The first part of the study examines the interdependencies between the US capital market and the UK and German markets by means of DCC-GARCH model [12, 13, 14, 31]. Table 1 shows the results of DCC-GARCH parameter estimation for the S&P500, FTSE100 and DAX indexes. The estimated parameters of the equation of conditional correlation and conditional variance for each index turned out to be statistically significant at the 5% significance level (excluding DAX autoregressive parameter). Statistical significance was also confirmed with respect to the parameters of conditional correlation equations.

The parameter v of the number of degrees of freedom of Student's t-distribution for the DCC-GARCH model was statistically significant, and the evaluation obtained at 9.624 indicates the existence of thick tails in the residual distribution. In addition, all parameter conditions are met since the sum of the parameters α1, β1and the sum of the parameters α and βare less than unity.

The conditional variance equations					
Parameter (Index)	Estimate	p value	Parameter (Index)	Estimate	p value
γ_0 (S&P500)	0.076	~0.000	α_1 (FTSE100)	0.101	~0.000
γ_1 (S&P500)	-0.056	0.002	β_1 (FTSE100)	0.891	~0.000
ω_1 (S&P500)	0.015	~0.000	γ_0 (DAX)	0.098	~0.000
α_1 (S&P500)	0.098	~0.000	γ_1 (DAX)	-0.002	0.902
β_1 (S&P500)	0.891	~0.000	ω_1 (DAX)	0.023	0.001
γ_0 (FTSE100)	0.055	0.002	α_1 (DAX)	0.093	~0.000
γ_1 (FTSE100)	-0.037	0.047	β_1 (DAX)	0.896	~0.000
ω_1 (FTSE100)	0.013	0.006	v	7.274	~0.000
The conditional correlation equation					
Parameter	Estimate	p value	Parameter	Estimate	p value
α	0.023	0.013	β	0.961	~0.000

(γ_0, γ_1 -- parameters of conditional mean equation, ω_1, α_1, β_1 - parameters of conditional variance equation, α, β - parameters of conditional correlation equation, v- degrees of freedom of t-distribution)

Table 1 The outcome of the estimation of the DCC-GARCH model

Using the DCC-GARCH model allows measuring the strength of interdependencies between the examined capital markets by means of a conditional correlation. Therefore, based on the DCC-GARCH parameter estima-

tion results, the conditional correlation values for the next pair of indices were determined. Analysis of changes in time of the value of conditional correlation allows determining trends in the formation of interdependence. Changes in the correlation values for a particular pair of indices indicate the strength of the interdependence between the two capital markets, as well as the upward or downward tendency of their evolution over time.

Figure 1 shows the conditional correlation between the US S&P500 index and the DAX and FTSE100 indices. The analysis of the correlation in the graphs shows that the relationship between the US capital market and the capital markets of the United Kingdom and Germany is high (the values range from 0.3 to 0.9). It can be said that due to the global financial crisis the level of interdependence increased significantly in the period from 2007 to the end of 2012. There has been a downward trend in the level of interdependence between selected markets since 2013, which indicates that the global financial crisis has been in decline. It is worth noting that there are slight differences in the interdependence between two pairs of indexes – the S&P500 and DAX pair and between S&P500 and FTSE100. This means that valuation of key European markets is heavily dependent on the US market and reactions ofEuropean markets to the US market are very similar.

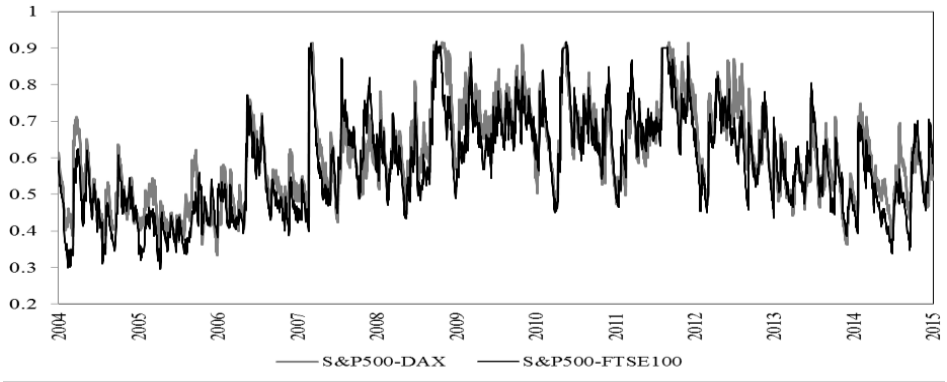


Figure 1 Interdependencies between the US S&P500 index and the major European indices DAX and FTSE100

3 Identification of a common, global factor for the US, the UK and German capital markets

In the second stage of the study, a cointegration analysis was performed for the conditional correlation processes obtained from the DCC-GARCH model [9, 11, 18, 32]. Before analyzingproblem of cointegration, the order of integration of conditional correlation processes was first investigated. Analysis of the degree of integration for interdependencies between pairs of selected capital markets was performed using the Phillips-Perron test[25,28]. The results of the test are contained in Table 2. For all three conditional correlation processes for pairs of indices (S&P500, FTSE100), (S&P500, DAX) and (FTSE100, DAX) calculated in successive steps of the test statistic values allowed us to indicate the order of integration equal to one I(1).This means that the investigated conditional correlation processes are non-stationary in variance. The non-stationarity of the processes may result from the nature of the connections between the capital markets selected in the analysis. Global information and shock events cause lasting changes in the interdependencies between the examined capital markets. The obtained results concerning the integration of conditional correlation processes lead to the formulation of a thesis assuming the existence of a similar mechanism of reaction of key EU capital markets to shocks from the US market.

Indices	**Time series**	**Test statistics**	**p-value**	**Order of integration**
S&P500, FTSE100	Levels	-1.384	0.154	I(1)
	Differences	-58.534	0.001	
S&P500, DAX	Levels	-1.263	0.191	I(1)
	Differences	-59.071	0.001	
FTSE100, DAX	Levels	-1.073	0.256	I(1)
	Differences	-54.863	0.001	

Table 2 Results of the integration analysis for the US, the UK and German capital markets

Determining the order of integration equal to one with the use of the Phillips and Perron test for all conditional correlation processes justifies the execution of cointegration analysis. Cointegration analysis was per-

formed on the basis of correlation processes for two pairs of indices, i.e.,the S&P500 and FTSE100pair, and the S&P500 and DAX pair. The analysis was conducted exclusively for two conditional correlation processes, since the identification of cointegration between these processes allows formulating a thesis of a similar long-term reaction mechanism for both European markets to global information and shock events. Consequently, this study seeks to ascertain whether conditional correlation processes for selected pairs of indices are linked by long-term relationships.

Max-Eigenvalue test		
Number of cointegrating vectors	Test statistic	p-value
None	48.178	~0.000
At most 1	1.725	0.222
Trace test		
Number of cointegrating vectors	Test statistic	p-value
None	46.453	~0.000
At most 1	1.725	0.222

Table 3 Results of the Johansen procedure

Correlation (Pair of indices)	**Parameter**	**Estimate**	**Standardized estimate**	**p-value**	**Factor Score Weights**
S&P500-DAX	α_1	1.000	0.973	-	0.018
S&P500-FTSE100	α_2	2.536	0.914	~0.00	0.016
FTSE100-DAX	α_3	1.416	0.955	~0.00	0.012

Table 4 Outcome of the confirmatory factor analysis

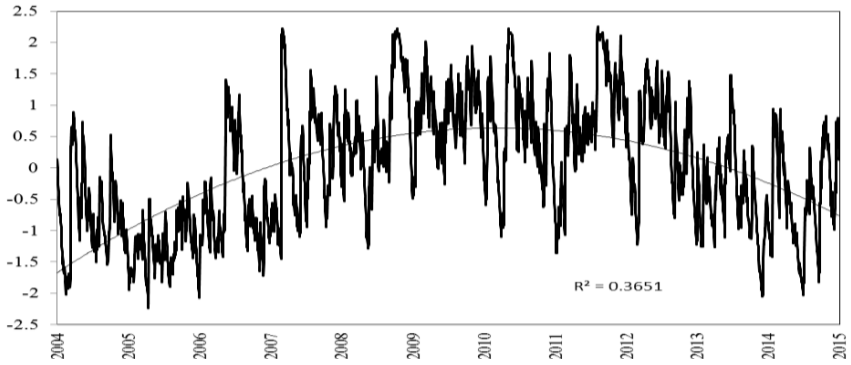


Figure 2 Common, global factor values identified based on a factor analysis

Johansen's cointegration analysis enabled us to identify the number of cointegrating relationships. Table 3 shows test results relating to the number of cointegrating vectors. Both the results of the maximum self-test and the trace test indicate the presence of one cointegrating vector. Identification of a single cointegrating vector means that the interdependence between the examined capital markets is linked by long-term relationships.The presence of cointegration justifies the possibility of separating a common, global factor for the interdependencies examined. The global factor, identified on the basis of confirmatory factor analysis, is a component that links the cross-correlation between the examined capital markets [3, 4, 8, 21, 26]. The results of the confirmatory factor analysis are presented in Table 4, where the prevalence of only one factor is assumed. All standardized evaluations are greater than 0.9, which indicates the correct co-generation of the factor by all three conditional correlation processes. In this way, one factor explaining the correlation between selected capital markets is established for the interdependencies of the markets. The values of the common global factor are derived from the sum of products of the values of Factors Score Weights (given in Table 4) and the values of the conditional correlation between the markets. Analysis of the value of a common, global factor makes it possible to evaluate changes in the whole pattern of relations between capital markets (see: Figure 2). Based on the obtained values of the common global factor, the parameters of the square trend model were estimated. The structural parameters for the

trend model proved to be statistically significant, the determination coefficient indicates a 36% explanation by the factor volatility model. Undoubtedly, the course of the square trend confirms the division into two stages of the formation of the interdependence system (or, in other words, the integration of the network ofrelationships) between the markets of the United States, the United Kingdom and Germany. The first period was associated with a gradual increase in the connections between the financial markets, the second with their weakening and the transition to a quiet period.

4 Conclusions

The paper presents the results of the analysis of the interdependence system in the capital markets of the United States, Germany, and the United Kingdom. A DCC-GARCH model was used to measure these interdependencies. The results of the analysis allowed us to assess the system of interdependence between these markets. In addition, a cointegration analysis was performed for conditional correlation processes. Identification of one cointegrating vector allowed us to conclude that the conditional correlation processes for selected pairs of indices are linked by a long-term relation. Determining the process of cointegration was a justification for identifying a common, global factor for the interdependence system. This factor was isolated on the basis of an affirmative factor analysis. A common, global factor is a component that links the interdependence of capital markets and presents changes in the whole system over time synthetically.

The interdependence study between the markets of the United States, the United Kingdom and Germany covered a period of global bull market precedingthe speculative bubble burst on the US real estate market, the biggest globalfinancial system volatility (2007-2010) in recent decadesand the period of gradual return of the markets to the so-called'normal functioning'. The results of the square trend estimate for the global factor have confirmed that market turbulence can be attributed to growing interdependencies between markets as a result of the global financial crisis. Then, according to the course of the trend, there was a slow stabilization of the situation on the capital markets and the transition of these markets to a period of calm.

References

[1] Arshanapalli, B., Doukasb, J., and Langc, L. H. P.: Pre and Post-October 1987 Stock Market Linkages Between U.S. and Asian Markets. *Pacific-Basin Finance Journal* **3(1)** (1995), 57-73.

[2] Balcerzak, A.P., Fałdziński, M., Meluzín, T., Pietrzak, M. B., and Zinecker, M.:Interdependence among Capital Markets of Germany, Poland and Baltic States. Application of DCC-GARCH Model. In:*Managing and Modeling Financial Risks. 8th International Scientific Conference Proceedings (Part I.)*(Čulík, M., ed.). VŠB-Technical University of Ostrava, Ostrava, 2016, 28-36.

[3] Balcerzak, A. P., and Pietrzak, M. B.: Quality of Human Capital in European Union in the Years 2004-2013. Application of Structural Equation Modeling. In: *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XVIII* (Reiff, M., and Gezik. P., eds.). Letra Interactive, Vratna, 2016, 7-12.

[4] Balcerzak, A. P., and Pietrzak, M. B.: Structural Equation Modeling in Evaluation of Technological Potential of European Union Countries in the Years 2008-2012. In: *The 10th Professor AleksanderZelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*. *Conference Proceedings* (Papież, M., and Śmiech, S., eds.). Foundation of the Cracow University of Economics, Cracow, 2016, 9-18.

[5] Balcerzak, A. P., and Rogalska, E.:Non-Keynesian Effects of Fiscal Consolidations in Central Europe in the Years 2000-2013.In:*Entrepreneurship, Business and Economics - Vol. 2. Proceedings of the 15th Eurasia Business and Economics Society*(Bilgin, M. H., and Danis, H., eds.). Springer International Publishing, 2016, 271-282.

[6] Baur, D.: Testing for Contagion – Mean and Volatility Contagion. *Journal of Multinational Financial Management* **13**(2003), 405-422.

[7] Billio, M., and Caporin, M.: Market Linkages, Variance Spillover and Correlation Stability: Empirical Evidences of Financial Contagion. *Computational Statistics & Data Analysis* **54(11)** (2010), 2443–2458.

[8] Bollen, K. A.: Structural Equations with Latent Variables. Wiley, 1989.

[9] Boswijk, H. P., and Doornik, J. A.: Identifying, Estimating and Testing Restricted Cointegrated Systems: An Overview. *StatisticaNeerlandica* **58(4)** (2004), 440–465.

[10] Corsetti, G., Pericoli, M., and Sbracia, M.: Some Contagion, Some Interdependence: More Pitfalls in Testing for Contagion. *Journal of International Money and Finance* **24** (2005), 1177-1199.

[11] Engle, R. F., and Granger, C. W. J.: Co-integration and Error Correction: Representation, Estimation and Testing. *Econometrica* **55(2)** (1987), 251–276.

[12] Engle, R. F.: Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business& Economic Statistics* **20** (2002), 339-350.
[13] Engle, R. F. :*Anticipating Correlations A New Paradigm for Risk Management.* Princeton University Press, Princeton and Oxford, 2009.
[14] Fałdziński, M., Balcerzak, A.P., Meluzín, T., Pietrzak, M. B., and Zinecker, M.: Cointegration of Interdependencies Among Capital Markets of Chosen Visegrad Countries and Germany. In. *34th International Conference Mathematical Methods in Economics MME 2016 Conference Proceedings*(Kocourek, A., and Vavrousek, M. eds.). Technical University of Liberec, Liberec, 2016, 189-194.
[15] Forbes K., and Rigobon R.:No Contagion, Only Interdependence: Measuring Stock Market Comovements, *Journal of Finance*, **57(5**) (2002), 2223-2261.
[16] Heryán, T., and Ziegelbauer, J.: Volatility of Yields of Government Bonds Among GIIPS Countries During the Sovereign Debt Crisis in the Euro Area. *Equilibrium.Quarterly Journal of Economics and Economic Policy* **11(1)** (2016), 61-74.
[17] Janus, J.: The Transmission Mechanism of Unconventional Monetary Policy. *OeconomiaCopernicana* **7(1)** (2016), 7-21.
[18] Johansen, S.: Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* **12(2–3)** (1988), 231-254.
[19] Kuc, M.: Social convergence in Nordic countries at regional level. *Equilibrium. Quarterly Journal of Economics and Economic Policy* **12(1)** (2017), 25-41.
[20] Lajtkepová, E.: Differences and Similarities in the Indebtedness of EU Member States after Last Financial Crisis. *OeconomiaCopernicana* **7(4)** (2016), 551-563.
[21] Loehlin, J. C.:*Latent variable models: An introduction to factor, path and structural analysis*. Erlbaum, 1987.
[22] Meluzín, T., and Zinecker, M.:Trends in IPOs: The Evidence from CEE Capital Markets. *Equilibrium. Quarterly Journal of Economics and Economic Policy* **11(2)** (2016), 327-341.
[23] Miklaszewicz, S.: Sovereign Debt Crisis of the Eurozone Countries. *OeconomiaCopernicana* **7(3)** (2016), 357-373.
[24] Murawska, A.: Differences in Unemployment Among Persons in a Special Situation on the Labour Market on the Example of Polish Provinces OeconomiaCopernicana **7(3)** (2016), 375-388
[25] Phillips, P., and Perron, P.: Testing for a Unit Root in Time Series Regression. *Biometrica*, **75(2)** (1988), 335-346.
[26] Pietrzak. M. B., and Balcerzak, A. P.:Socio-economic Sustainability in Poland. SEM Analysis atRegional Level. In:*16th International Scientific Conference Globalization and Its Socio-Economic Consequences. University of Zilina, The Faculty of Operation and Economics of Transport and Communication, Department of Economics, 5th – 6th October 2016. (Part IV.)*(Kliestik, T., ed.).University of Zilina, Zilina, 2016, 1704-1712.
[27] Pietrzak, M. B., and Balcerzak, A. P.: A Spatial SAR Model in Evaluating Influence of Entrepreneurship and Investments on Unemployment in Poland. In: *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XVIII* (Reiff, M., and Gezik, P., eds.). Letra Interactive, Vratna, 2016, 303-308.
[28] Pietrzak, M. B., Fałdziński, M., Balcerzak, A. P., Meluzín, T., and Zinecker, M.:Short-term Shocks and Long-term Relationships of Interdependencies Among Central European Capital Markets. *Economics & Sociology* **10(1)** (2017), 61-77.
[29] Pritsker M.: The Channel for Financial Contagion. In: *International Financial Contagion: How It Spreads and How It Can Stopped* (Claessens, S., and Forbes, K. J., eds.). Kluwer Academic Publishers, Boston/Dordrecht/London, 2001.
[30] Stefaniak-Kopoboru, J., and Kuczewska, J.: Export Specialization in Services of the Visegrad Countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy* **11(2)** (2016), 265-284.
[31] Szumilo, N., Bienert, S., Łaszkiewicz, E., Pietrzak, M. B., and Balcerzak, A. P.: The Real alternative? A comparison of German real estate returns with bonds and stocks. *Journal of Property Investment & Finance*, **35**, 2017.
[32] Zinecker, M., Balcerzak, A. P.,Fałdziński, M., Meluzín, T., and Pietrzak, M. B.: Application of DCC-GARCH Model for Analysis of Interrelations Among Capital Markets of Poland, Czech Republic and Germany. In *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XVIII*(Reiff, M., and Gezik, P., eds.). Letra Interactive,Vratna. 2016, 416-421.
[33] Żelazny, R., and Pietrucha, J.:Measuring innovation and institution: the creative economy index. *Equilibrium. Quarterly Journal of Economics and Economic Policy* **12(1)** (2017), 43-62.

Efficient algorithms for X-simplicity and weak $(\mathbf{X}, \lambda)$-robustness of fuzzy matrices

Ján Plavka [1]

Abstract. A fuzzy (max-min) matrix A is said to have **X**-simple image eigenspace if any eigenvector x belonging to the interval **X** is the unique solution of the max-min linear system with matrix of the system A and right-hand vector x and the unique solution is in **X**. The main result of this contribution is to present efficient algorithms for **X**-simplicity and to generalize the results to interval version of weak $(\mathbf{X}, \lambda)$-robustness.

Keywords: max-min algebra, interval, weakly stable, eigenspace, simple image set

JEL classification: C44
AMS classification: 15A80, 15A18, 08A72

1 Introduction

Max-min algebra (the addition and the multiplication are formally replaced by operations of maximum and minimum) can be used in a range of practical problems related to scheduling, optimization, modeling of fuzzy discrete dynamic systems, graph theory, knowledge engineering, cluster analysis and fuzzy systems.

In the max-min algebra, sometimes also called the "fuzzy algebra" [3, 4, 8, 10, 11], the arithmetical operations $a \oplus b := \max(a, b)$ and $a \otimes b := \min(a, b)$ are defined over a linearly ordered set. As usual, the two arithmetical operations are naturally extended to matrices and vectors.

Max-min algebra was historically motivated by multi-machine interaction processes ([1, 2, 3, 5, 6, 7, 9, 15]. In these processes we have n machines which work in stages, and in the algebraic model of their interactive work, entry $x_i^{(k)}$ of a vector $x^{(k)} \in \mathbb{B}(n, n)$ where $i \in \{1, \ldots, n\}$ and $\mathbb{B}$ is an idempotent semiring, represents the state of machine i after some stage k, and the entry a_{ij} of a matrix $A \in \mathbb{B}(n, n)$, where $i, j \in \{1, \ldots, n\}$, encodes the influence of the work of machine j in the previous stage on the work of machine i in the current stage. Summing up all the influence effects multiplied by the results of previous stages, we have $x_i^{(k+1)} = \bigoplus_j a_{ij} \otimes x_j^{(k)}$. In the case of $\oplus = \max$ this "summation" is often interpreted as waiting till all the processes are finished and all the necessary influence constraints are satisfied.

The orbit x, $A \otimes x, \ldots, A^k \otimes x$, where $A^k = A \otimes \ldots \otimes A$, represents the evolution of such a process. Regarding the orbits, one wishes to know the set of starting vectors from which a given objective can be achieved. One of the most natural objectives in max-min algebra, where the ultimate periodicity of the orbits often occurs, is to arrive at an eigenvector. The set of starting vectors from which one reaches an eigenvector of A after a finite number of stages, is called attraction set of A (see [1]). In general, attraction set contains the set of all eigenvectors, but it can be also as big as the whole space. This leads us, in turn, to another question: in which case is attraction set precisely the same as the set of all eigenvectors? Matrices with this property are called weakly robust [12] or weakly stable [1]. In terms of the systems $A \otimes x = b$, weak robustness is equivalent to the following condition: every eigenvector y belongs to the *simple image set* of A, that is, for every eigenvector y, the system $A \otimes x = y$ has unique solution $x = y$ (see [11]).

In the present paper, we consider an interval version of this condition. Namely, we describe matrices $A \in \mathbb{B}(n, n)$ such that for any eigenvector y belonging to an interval $\boldsymbol{X} = [\underline{x}, \overline{x}] := \{x \in \mathbb{B}(n);\ \underline{x} \leq x \leq \overline{x}\}$ the system $A \otimes x = y$ has a unique solution $x = y$ in $\boldsymbol{X}$. This is what we mean by saying that "A has $\boldsymbol{X}$-simple image eigenspace". It was shown in [13] that A has $\boldsymbol{X}$-simple image eigenspace if and only if it satisfies a nontrivial combinatorial criterion, which makes use of threshold digraphs and to which we refer as "$\boldsymbol{X}$-conformity" (see Definition 6).

Let us now give more details on the organization of the paper and on the results obtained there. The next section will be occupied by some definitions and notation of the max-min algebra, leading to the discussion of weak $\boldsymbol{X}$-robustness and $\boldsymbol{X}$-simple image eigenvectors. Section 3 is devoted to the main results of the paper which characterizes matrices with $\boldsymbol{X}$-simple image eigenspaces and weak $(\mathbf{X}, \lambda)$-robustness.

[1]Technical University in Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, Jan.Plavka@tuke.sk

2 Preliminaries

2.1 Max-min algebra and associated digraphs

Let us denote the set of all natural numbers by $\mathbb{N}$. Let $(\mathbb{B}, \leq)$ be a bounded linearly ordered set with the least element in $\mathbb{B}$ denoted by O and the greatest one by I.

A max-min algebra is a set $\mathbb{B}$ equipped with two binary operations $\oplus = \max$ and $\otimes = \min$, called addition and multiplication, such that $(\mathbb{B}, \oplus)$ is a commutative monoid with identity element O, $(\mathbb{B}, \otimes)$ is a monoid with identity element I, multiplication left and right distributes over addition and multiplication by O annihilates $\mathbb{B}$.

We will use the notations N and M for the sets of natural numbers not exceeding n and m, respectively, i.e., $N = \{1, 2, \ldots, n\}$ and $M = \{1, 2, \ldots, m\}$. The set of $n \times m$ matrices over $\mathbb{B}$ is denoted by $\mathbb{B}(n, m)$, and the set of $n \times 1$ vectors over $\mathbb{B}$ is denoted by $\mathbb{B}(n)$. If each entry of a matrix $A \in \mathbb{B}(n, n)$ (a vector $x \in \mathbb{B}(n)$) is equal to O we shall denote it as $A = O$ ($x = O$).

Let $x = (x_1, \ldots, x_n) \in \mathbb{B}(n)$ and $y = (y_1, \ldots, y_n) \in \mathbb{B}(n)$ be vectors. We write $x \leq y$ ($x < y$) if $x_i \leq y_i$ ($x_i < y_i$) holds for each $i \in N$.

For a matrix $A \in \mathbb{B}(n, n)$ the symbol $G(A) = (N, E)$ stands for a complete, arc-weighted digraph associated with A, i.e., the node set of $G(A)$ is N, and the weight (capacity) of any arc (i, j) is $a_{ij} \geq O$. For given $h \in \mathbb{B}$, the *threshold digraph* $G(A, h)$ is the digraph with the node set N and with the arc set $E = \{(i, j);\ i, j \in N,\ a_{ij} \geq h\}$. A path in the digraph $G(A) = (N, E)$ is a sequence of nodes $p = (i_1, \ldots, i_{k+1})$ such that $(i_j, i_{j+1}) \in E$ for $j = 1, \ldots, k$. The number k is the length of the path p and is denoted by $l(p)$. If $i_1 = i_{k+1}$, then p is called a cycle and it is called an elementary cycle if moreover $i_j \neq i_m$ for $j, m = 1, \ldots, k$.

A matrix A is called *generalized Hamiltonian permutation* if all nonzero entries of A lie on a Hamiltonian cycle (the threshold digraph $G(A, h),\ h = \min\{a_{ij};\ a_{ij} > O,\ i, j \in N\}$ is elementary cycle containing all nodes).

2.2 Orbits, eigenvectors and weak λ-robustness

For $A \in \mathbb{B}(n, n)$ and $x \in \mathbb{B}(n)$, the orbit $O(A, x)$ of $x = x^{(0)}$ generated by A is the sequence of vectors $x^{(0)}, x^{(1)}, \ldots, x^{(n)}, \ldots$, where $x^{(r)} = A^r \otimes x^{(0)}$ for each $r \in \mathbb{N}$.

The operations $\max, \min$ are idempotent, so no new numbers are created in the process of generating of an orbit. Therefore any orbit contains only a finite number of different vectors. It follows that any orbit starts repeating itself after some time, in other words, it is ultimately periodic. The same holds for the power sequence $(A^k; k \in \mathbb{N})$.

We are interested in the case when the ultimate period is 1, or in other words, when the orbit is ultimately stable. Note that in this case the ultimate vector of the orbit necessarily satisfies $A \otimes x = x$. This is the main reason why in this paper by eigenvectors we mean fixed points. (Also observe that if x is not a fixed point but a more general eigenvector satisfying $A \otimes x = \lambda \otimes x$, then $A \otimes x$ is already a fixed point due to the idempotency of multiplication.)

For a given matrix $A \in \mathbb{B}(n, n)$, the number $\lambda \in \mathbb{B}$ and the n-tuple $x \in \mathbb{B}(n)$ are the so-called *eigenvalue* of A and *eigenvector* of A, respectively, if $A \otimes x = \lambda \otimes x$.

The *eigenspace* $V(A, \lambda)$ is defined as the set of all eigenvectors of A with associated eigenvalue λ, i.e., $V(A, \lambda) = \{x \in \mathbb{B}(n);\ A \otimes x = \lambda \otimes x\}$.

Formally we can define the attraction set $\mathrm{attr}(A, \lambda)$ as follows

$$\mathrm{attr}(A, \lambda) = \{x \in \mathbb{B}(n);\ O(A, x) \cap V(A, \lambda) \neq \emptyset\}.$$

In case $\lambda = I$ let us denote $V(A, I)$ and $\mathrm{attr}(A, \lambda)$ by abbreviation $V(A)$ and $\mathrm{attr}(A)$, respectively.

Definition 1. Let $A \in \mathbb{B}(n, n), \lambda \in \mathbb{B}$ be given. Then A is called weakly λ-robust, if $\mathrm{attr}(A, \lambda) = V(A, \lambda)$.

Observe that in general $V(A, \lambda) \subseteq \mathrm{attr}(A, \lambda) \subseteq \mathbb{B}^n$. The matrices for which $\mathrm{attr}(A, \lambda) = \mathbb{B}^n$ are called (strongly) robust or (strongly) stable, as opposed to weakly robust (weakly stable).

Theorem 1. [12] *Let $A \in \mathbb{B}(n, n), \lambda \in \mathbb{B}$ be given. Then A is weakly λ-robust if and only if* $(\forall x \in \mathbb{B}(n))[A \otimes x \in V(A, \lambda) \Rightarrow x \in V(A, \lambda)]$.

Let us conclude this section with recalling some information on 1) the greatest eigenvector and 2) constant eigenvectors in max-min algebra.

Let $A = (a_{ij}) \in \mathbb{B}(n,n)$, $\lambda \in \mathbb{B}$ be given and define the greatest eigenvector $x^{\oplus}(A,\lambda)$ corresponding to a matrix A and λ as

$$x^{\oplus}(A,\lambda) = \bigoplus_{x \in V(A,\lambda)} x.$$

In case $\lambda = I$ let us denote $x^{\oplus}(A,I)$ by abbreviation $x^{\oplus}(A)$.

In [14] it was stated that the greatest eigenvector $x^{\oplus}(A,\lambda)$ exists for every matrix A whereby its entries are given by the efficient formula presented in the next theorem.

Theorem 2. [14] *Let $A \in B(n,n)$, $\lambda \in \mathbb{B}$ be given and $r_i = \bigoplus_{j \in N} a_{ji}$. Then*

$$x_i^{\oplus}(A,\lambda) = \begin{cases} I, & \text{if } \lambda \oplus r_i(A) \leq \lambda \otimes (A^n \otimes I^{\star})_i, \\ \lambda \otimes (A^n \otimes I^{\star})_i, & \text{otherwise.} \end{cases}$$

Next, denote

$$c(A) = \bigotimes_{i \in N} \bigoplus_{j \in N} a_{ij}, \quad c^*(A) = (c(A), \ldots, c(A))^T \in \mathbb{B}(n).$$

It can be checked that $A \otimes c^*(A) = c^*(A)$, so $c^*(A)$ is a constant eigenvector of A. As $x^{\oplus}(A)$ is the greatest eigenvector of A, we have $c^*(A) \leq x^{\oplus}(A)$.

Theorem 3. [12] *Let $A \neq O$ and $\lambda > O$. If A is weakly λ-robust then A is a permutation matrix.*

Now, let us suppose that $A = (a_{ij}) \in B(n,n)$ is a permutation matrix and $\lambda \in B$. Then the digraph $G(A, c(A)$ is the set of Hamiltonian cycles, say $c_i = (k_1^i, \ldots, k_{l_i}^i)$ for $i \in S = \{1, \ldots, s\}$. Without loss of generality the matrix A can be considered in block-diagonal form (denoted by $A = (A_1, \ldots, A_s)$)

$$A = \begin{pmatrix} A_1 & O & \ldots & O \\ O & A_2 & \ldots & O \\ \vdots & & & \\ O & O & \ldots & A_s \end{pmatrix}, \tag{1}$$

where each submatrix A_i is generalized Hamiltonian permutation and corresponds to the cycle $c_i = (k_1^i, \ldots, k_{l_i}^i)$.

Theorem 4. [12] *Let $A \in B(n,n)$, $A \neq O$, $A = (A_1, \ldots, A_s)$, $s \geq 2$ be a block-diagonal permutation matrix and $\lambda \in B$. Then A is weakly λ-robust if and only if $(\forall i \in S)[\lambda < c(A_i) \ \vee \ \lambda = c(A_i) = m(A_i)]$.*

Algorithm **Weak λ-Robustness**

Input. $A = (a_{ij})$, $\lambda \in B$.
Output. 'yes' in variable wr if A is weakly λ-robust; 'no' in wr otherwise.
begin

if $A = O$ and $\lambda = O$ **then** wr:='yes' **else** wr:='no'

if $A \neq O$ and $\lambda > O$ and $A \approx (A_1, \ldots, A_s)$ and

$(\forall i \in \{1, \ldots, s\})[\lambda < c(A_i) \ \vee \ \lambda = c(A_i) = m(A_i)]$ **then** wr:='yes' **else** wr:='no'

end

Theorem 5. [12] *Let A be a fuzzy matrix and $\lambda \in B$. The algorithm* **Weak λ-Robustness** *correctly decides whether a matrix A is weakly λ-robust in $O(n^2)$ arithmetic operations.*

2.3 Weak $\boldsymbol{X}$-robustness and $\boldsymbol{X}$-simplicity

In this subsection we consider an interval extension of weak robustness and its connection to $\boldsymbol{X}$-simplicity. We remind that throughout the paper,

$$\boldsymbol{X} = [\underline{x}, \overline{x}] = \{x \colon \underline{x} \leq x \leq \overline{x}, \}, \text{ where } \underline{x}, \overline{x} \in \mathbb{B}(n).$$

Consider the following interval extension of weak $\boldsymbol{X}$-robustness.

Definition 2. $A \in \mathbb{B}(n,n)$ is called weakly $\boldsymbol{X}$-robust if $\mathrm{attr}(A) \cap \boldsymbol{X} \subseteq V(A)$.

The notion of $\boldsymbol{X}$-simplicity is related to the concept of simple image set [1]: by definition, this is the set of vectors b such that the system $A \otimes x = b$ has a unique solution, which is usually denoted by $|S(A,b)| = 1$ ($S(A,b)$ standing for the solution set of $A \otimes x = b$). If the only solution of the system $A \otimes x = b$ is $x = b$, then b is called a *simple image eigenvector.*

If $\boldsymbol{X} = \mathbb{B}$ then the notion of weak robustness can be described in terms of simple image eigenvectors:

Proposition 6. [13] *Let $A \in \mathbb{B}(n,n)$. The following are equivalent:*
(i) A is weakly robust;
(ii) $(\forall x \in V(A))[|S(A,x)| = 1]$;
(iii)Each $x \in V(A)$ is a simple image eigenvector.

This motivates us to consider an interval version of simple image eigenvectors.

Definition 3. Let $A = (a_{ij}) \in \mathbb{B}(n,n)$.
(i) An eigenvector $x \in V(A) \cap \boldsymbol{X}$ is called an $\boldsymbol{X}$-simple image eigenvector if x is the unique solution of the equation $A \otimes y = x$ in interval $\boldsymbol{X}$.
(ii) Matrix A is said to have $\boldsymbol{X}$-simple image eigenspace if any $x \in V(A) \cap \boldsymbol{X}$ is an $\boldsymbol{X}$-simple image eigenvector.

Definition 4. Let $A \in \mathbb{B}(n,n)$ and $\boldsymbol{X} = [\underline{x}, \overline{x}] \subseteq \mathbb{B}(n)$ be given. We say that $\boldsymbol{X}$ is invariant under A if $x \in \boldsymbol{X}$ implies $A \otimes x \in \boldsymbol{X}$.

Theorem 7. [13] *Let $A \in \mathbb{B}(n,n)$ be a matrix and $\boldsymbol{X} = [\underline{x}, \overline{x}] \subseteq \mathbb{B}(n)$ be an interval vector.*
(i) If A is weakly $\boldsymbol{X}$-robust then A has $\boldsymbol{X}$-simple image eigenspace.
(ii) If A has $\boldsymbol{X}$-simple image eigenspace and if $\boldsymbol{X}$ is invariant under A then A is weakly $\boldsymbol{X}$-robust.

Proposition 8. [13] *$\boldsymbol{X}$ is invariant under A if and only if $A \otimes \underline{x} \geq \underline{x}$ and $A \otimes \overline{x} \leq \overline{x}$.*

Thus the $\boldsymbol{X}$-simplicity is a necessary condition for weak $\boldsymbol{X}$-robustness. It is also sufficient if the interval $\boldsymbol{X}$ is invariant under A, i.e., $\underline{x} \leq A \otimes \underline{x}$ and $A \otimes \overline{x} \leq \overline{x}$.

3 Interval $\boldsymbol{X}$-simplicity

The purpose of this section is to define the condition for matrix A which will ensure that each eigenvector $x \in V(A) \cap \boldsymbol{X}$ is an $\boldsymbol{X}$-simple image eigenvector and to deal with matrices with interval elements. Sufficient and necessary conditions for an interval matrix which have $\boldsymbol{X}$-simple eigenspace will be proved. In addition we introduce a polynomial algorithm to check the $\boldsymbol{X}$-simplicity of interval fuzzy matrices.

Definition 5. Let $A \in \mathbb{B}(n,n)$ be a matrix. A is called a *generalized level α-permutation matrix* (abrr. level α-permutation) if all entries greater than or equal to α of A lie on disjoint elementary cycles covering all the nodes. In other words, the threshold digraph $G(A,\alpha)$ is the set of disjoint elementary cycle containing all nodes.

Remark 1. If $A = (a_{ij}) \in \mathbb{B}(n,n)$ is a level $c(A)$-permutation matrix, $(i_1,\ldots,i_n)$ is a permutation of N such that $a_{i_j i_{j+1}} \geq c(A)$ and $(i_1,\ldots,i_n) = (i_1^1,\ldots,i_{s_1}^1)\ldots(i_1^k,\ldots,i_{s_k}^k)$ ($c_u = (i_1^u,\ldots,i_{s_u}^u)$ is an elementary cycle in digraph $G(A,c(A))$, $u = 1,\ldots,k$), then $x_v^{\oplus}(A) = \min\limits_{(k,l)\in c_u} a_{kl}$ for all $v \in c_u$ (see [12]).

Lemma 9. [13] Let $A = (a_{ij}) \in \mathbb{B}(n,n)$ be a matrix, $\boldsymbol{X} = [\underline{x}, \overline{x}] \in \mathbb{B}(n)$ be an interval vector and $\underline{x} < c^*(A) \leq \overline{x}$. If A has $\boldsymbol{X}$-simple image eigenspace then A is level $c(A)$-permutation.

Definition 6. Let $\boldsymbol{X} = [\underline{x}, \overline{x}] \subseteq \mathbb{B}(n)$ be an interval vector such that $\underline{x} < c^*(A) \leq \overline{x}$ and $A = (a_{ij}) \in \mathbb{B}(n,n)$ be a level $c(A)$-permutation matrix, $(i_1,\ldots,i_n)$ be a permutation of N such that $a_{i_j i_{j+1}} \geq c(A)$ and $(i_1,\ldots,i_n) = (i_1^1,\ldots,i_{s_1}^1)\ldots(i_1^k,\ldots,i_{s_k}^k)$ ($c_u = (i_1^u,\ldots,i_{s_u}^u)$ be an elementary cycle in digraph $G(A,c(A))$, $u = 1,\ldots,k$). Then vectors $e_{\underline{x}} = (e_1,\ldots,e_n)^T$ and $f_{\overline{x}}(A) = (f_1,\ldots,f_n)^T$ are called $\underline{x}$-vector of A and $\overline{x}$-vector of A if $e_i = \max\limits_{v\in c_u} \underline{x}_v$ and $f_i = \min\limits_{v\in c_u} \overline{x}_v \otimes x_v^{\oplus}(A)$, respectively, for $i \in c_u$, $u \in \{1,\ldots,k\}$.
Matrix A is called $\boldsymbol{X}$-conforming if
(i) $\underline{x}_{i_{j+1}} < e_{i_{j+1}} \Rightarrow a_{i_j k} < e_{i_j}$ for $k \neq i_{j+1},\ k \in N$
(ii) $\underline{x}_{i_{j+1}} = e_{i_{j+1}} \Rightarrow a_{i_j k} \leq e_{i_j}$ for $k \neq i_{j+1},\ k \in N$
(iii)$a_{i_j i_{j+1}} = \min\limits_{(k,l)\in c_u} a_{kl} = x_{i_{j+1}}^{\oplus}(A) = f_{i_{j+1}} \Rightarrow \overline{x}_{i_{j+1}} \leq x_{i_{j+1}}^{\oplus}(A)$.

Notice that $e_{i_j} = e_{i_{j+1}}$ and $f_{i_j} = f_{i_{j+1}}$ by definition of $e_{\underline{x}}$ and $f_{\overline{x}}$ (nodes i_j, i_{j+1} are lying in the same cycle c_u). Notation $(k,l) \in c_u$ means that the edge (k,l) is lying in c_u.

Theorem 10. [13] *Let $A = (a_{ij}) \in \mathbb{B}(n,n)$ be a matrix, $\boldsymbol{X} = [\underline{x}, \overline{x}] \in \mathbb{B}(n)$ be an interval vector and $\underline{x} < c^*(A) \leq \overline{x}$. Then A has $\boldsymbol{X}$-simple image eigenspace if and only if A is an $\boldsymbol{X}$-conforming matrix.*

Algorithm **X-comforming**

Input. $A = (a_{ij}) \in \mathbb{B}(n,n)$, $\underline{x}, \overline{x} \in \mathbb{B}(n)$.
Output. 'yes' in variable $Xcom$ if A is $\boldsymbol{X}$-comforming; 'no' in $Xcom$ otherwise.
begin

if A is not level $c(A)$-permutation **then** wr:='no' **else compute** $e_{\underline{x}}$, $f_{\overline{x}}$ and $x^{\oplus}(A)$;

if $(i_1, \ldots, i_n) = (i_1^1, \ldots, i_{s_1}^1) \ldots (i_1^k, \ldots, i_{s_k}^k)$ is a permutation of N such that $a_{i_j i_{j+1}} \geq c(A)$ and

(i) $\underline{x}_{i_{j+1}} < e_{i_{j+1}} \Rightarrow a_{i_j k} < e_{i_j}$ for $k \neq i_{j+1},\ k \in N$

(ii) $\underline{x}_{i_{j+1}} = e_{i_{j+1}} \Rightarrow a_{i_j k} \leq e_{i_j}$ for $k \neq i_{j+1},\ k \in N$

(iii) $a_{i_j i_{j+1}} = \min\limits_{(k,l) \in c_u} a_{kl} = x^{\oplus}_{i_{j+1}}(A) = f_{i_{j+1}} \Rightarrow \overline{x}_{i_{j+1}} \leq x^{\oplus}_{i_{j+1}}(A)$

then $Xcom$:='yes' **else** $Xcom$:='no'

end

Notice that the complexity of checking that a given matrix A has $\boldsymbol{X}$-simple image eigenspace requires $O(n^2 \log n)$ arithmetic operations. It is based on the fact that we need $O(n^2)$ operations to check parts (i) and (ii) and $O(n^2 \log n)$ operations to compute the greatest eigenvector $x^{\oplus}(A)$ ([14]).

3.1 Weak $(\boldsymbol{X}, \lambda)$-robustness

In this subsection we consider an extension of weak $\boldsymbol{X}$-robustness and its connection to $(\boldsymbol{X}, \lambda)$-simplicity.

If $\boldsymbol{X} = \mathbb{B}$ then the notion of weak robustness can be described in terms of simple image eigenvectors:

Proposition 11. *Let $A \in \mathbb{B}(n,n)$, $\lambda \in \mathbb{B}$. The following are equivalent:*
(i) A is weakly λ-robust;
(ii) $(\forall x \in V(A,\lambda))[S(A, \lambda \otimes x) = \{y \in \mathbb{B}(n); \lambda \otimes y = \lambda \otimes x\}]$;

Proof. Suppose that A is weakly λ-robust and there is $x \in V(A,\lambda)$ s.t. $S(A, \lambda \otimes x) \neq \{y \in \mathbb{B}(n); \lambda \otimes y = \lambda \otimes x\}]$, i.e., there is $y \in (S(A, \lambda \otimes x)$ such that $A \otimes y = \lambda \otimes x \neq \lambda \otimes y$. By Theorem 1 we get $A \otimes (A \otimes y) = A \otimes (\lambda \otimes x) = \lambda \otimes x$ and $A \otimes y = \lambda \otimes x \neq \lambda \otimes y$, this is a contradiction.
The converse implication trivially follows. □

The notion of $(\boldsymbol{X}, \lambda)$-simplicity, which we will introduce next, is related to the concept of $\boldsymbol{X}$-simplicity. If the solution set $S(A,b)$ of the system $A \otimes x = \lambda \otimes b$ is $\{y \in \mathbb{B}(n); \lambda \otimes y = \lambda \otimes b\}$, then b is called an *simple image λ-eigenvector.*

Notice that if $\lambda = I$, then the notions simple image λ-eigenvector and simple image eigenvector are equivalent.

Consider the following interval extension of weak λ-robustness.

Definition 7. Let $A \in \mathbb{B}(n,n), \lambda \in \mathbb{B}$ be given. A is called weakly $(\boldsymbol{X}, \lambda)$-robust if $\mathrm{attr}(A, \lambda) \cap \boldsymbol{X} \subseteq V(A, \lambda)$.

The next assertion is an interval version of Proposition 11.

Proposition 12. *Let $A \in \mathbb{B}(n,n), \lambda \in \mathbb{B}$. The following are equivalent:*
(i) A is weakly $(\boldsymbol{X}, \lambda)$-robust;
(ii) $(\forall x \in V(A,\lambda) \cap \boldsymbol{X})[S(A, \lambda \otimes x) \cap \boldsymbol{X} = \{y \in \mathbb{B}(n); \lambda \otimes y = \lambda \otimes x\} \cap \boldsymbol{X}]$.

This motivates us to consider a generalized interval version of $\boldsymbol{X}$-simple image eigenvectors.

Definition 8. Let $A = (a_{ij}) \in \mathbb{B}(n,n), \lambda \in \mathbb{B}$.
(i) An eigenvector $x \in V(A,\lambda) \cap \boldsymbol{X}$ is called an $\boldsymbol{X}$-simple image λ-eigenvector if the set solutions $S(A, \lambda \otimes x)$ of the equation $A \otimes y = \lambda \otimes x$ is equal to $\{y \in \boldsymbol{X}; \lambda \otimes y = \lambda \otimes x\}$.
(ii) Matrix A is said to have $(\boldsymbol{X}, \lambda)$-simple image eigenspace if any $x \in V(A,\lambda) \cap \boldsymbol{X}$ is an $\boldsymbol{X}$-simple image λ-eigenvector.

Theorem 13. *Let $A \in \mathbb{B}(n,n)$, $\boldsymbol{X} = [\underline{x}, \overline{x}]$, $\lambda \in \mathbb{B}$ be given and $\lambda \geq \max\limits_{i \in N} \underline{x}_i$.*

(i) If A is weakly $(\boldsymbol{X}, \lambda)$-robust then A has $(\boldsymbol{X}, \lambda)$-simple image eigenspace.
(ii) If A has $(\boldsymbol{X}, \lambda)$-simple image eigenspace and if $\boldsymbol{X}$ is invariant under A then A is weakly $(\boldsymbol{X}, \lambda)$-robust.

Proof. (i) Suppose that A is weakly $(\boldsymbol{X}, \lambda)$-robust and $x \in V(A) \cap \boldsymbol{X}$. If the system $A \otimes y = \lambda \otimes x$ has a solution $y \in \boldsymbol{X}$ such that $\lambda \otimes y \neq \lambda \otimes x$, then y is not an eigenvector but belongs to $\mathrm{attr}(A) \cap \boldsymbol{X}$, which contradicts the weak $(\boldsymbol{X}, \lambda)$-robustness.

(ii) Assume that A has $(\boldsymbol{X}, \lambda)$-simple image eigenspace and x is an arbitrary element of $\mathrm{attr}(A, \lambda) \cap \boldsymbol{X}$. As $\boldsymbol{X}$ is invariant under A, we have that $A^k \otimes x \in \boldsymbol{X}$ for all k. Then $A^k \otimes x \in V(A, \lambda)$ for some k implies $A^{k-1} \otimes x = A^k \otimes x \in V(A, \lambda)$,..., $x \in V(A, \lambda)$ because of $\lambda \geq \max\limits_{i \in N} \underline{x}_i$. □

If $\lambda \geq \max\limits_{i \in N} \underline{x}_i$, then the $\boldsymbol{X}$-simplicity is a necessary condition for weak $\boldsymbol{X}$-robustness. It is also sufficient if the interval $\boldsymbol{X}$ is invariant under A, i.e., $(\lambda \otimes \underline{x} =)\, \underline{x} \leq A \otimes \underline{x}$ and $A \otimes \overline{x} \leq \overline{x}$.

References

[1] Butkovič, P.: Max-linear Systems: Theory and Algorithms, *Springer Monographs in Mathematics*, Springer-Verlag 2010.

[2] Cuninghame-Green, R. A.: Minimax Algebra, *Lecture Notes in Economics and Mathematical Systems* vol. 166, Springer, Berlin, 1979.

[3] Gavalec, M.: Periodicity in Extremal Algebra, Gaudeamus, Hradec Králové 2004.

[4] Molnárová, M., Myšková, H. and Plavka, J.: The robustness of interval fuzzy matrices, *Linear Algebra and Its Applications* 438(8) (2013) 3350–3364.

[5] Molnárová, M.: Generalized matrix period in max-plus algebra, *Linear Algebra and Its Applications* 404 (2005) 345–366.

[6] Molnárová, M.: Computational complexity of Nachtigall's representation, *Optimization* 52(1) (2003) 93–104.

[7] Molnárová, M. and Pribiš, J.: Matrix period in max-algebra, *Discrete Applied Mathematics* 103(1-3) (2000) 167–175.

[8] Myšková, H.: Interval max-plus matrix equations, *Linear Algebra and Its Applications* 492 (2016) 111–127.

[9] Myšková, H.: An iterative algorithm for testing solvability of max-min systems, *Kybernetika* 48(5) (2012), 879–889.

[10] Myšková, H.: Max-min interval systems of linear equations with bounded solution, *Kybernetika* 48(2) (2012), 299-308.

[11] Plavka J., and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *Discrete Applied Mathematics* 159(5) (2011), 381–388.

[12] Plavka, J.: On the weak robustness of fuzzy matrices, *Kybernetika* 49(1) (2013), 128–140.

[13] Plavka, J., and Sergeev, S.: Characterizing matrices with X-simple image eigenspace in max-min semiring, *Kybernetika* 52(4) (2016), 497–513.

[14] Tan, Yi-Jia: On the eigenproblem of matrices over distributive lattices, *Linear Algebra and its Applications* 374 (2003), 87–106.

[15] Zimmernann, K.: Extremální algebra (in Czech), *Ekon. ústav ČSAV* Praha, 1976.

Using the Sweep Algorithm for decomposing a set of vertices and subsequent solution of the traveling salesman problem in decomposed subsets

Marek Pomp[1], Petr Kozel[2], Šárka Michalcová[3], Lucie Orlíková[4]

Abstract. The aim of the Traveling Salesman Problem is to find an optimal route, which passes through all vertices of the given graph just once (except for the initial vertex) and that is the one with a minimum length. Basic requirement for solution is that service vehicle capacity holds for all requirements, which are placed at vertices of graph (it could be a collection or distribution). However, it is a common practice that capacity of service vehicle is less than all requirements, which must be satisfied. One of basic ways for troubleshooting is a decomposition of the input set of vertices into several subsets so that the sum of capacity requirements in individual subsets does not exceed the capacity of service vehicle. In these subsets, the optimal route can be found by some of the accessible ways: an exact solution in case of small-scale tasks, heuristics in case of large-scale tasks. The core of the present paper deals with the use of the concept of the Sweep Algorithm for decomposing input set of vertices and subsequent determining optimal vehicle routes using an exact algorithm. Computational experiments that will be presented in this paper were realized in conditions of the optimal route planning for service vehicles, which ensure the municipal waste collection of a private company.

Keywords: Sweep Algorithm, Cluster-First Route-Second, ArcGIS, "R",vehicle routing problems, Hamiltonian circuit, municipal waste collection.

JEL classification: C61
AMS classification: 90C05

1 Introduction

Service vehicle route planning can be considered as a very important activity of subject that is dedicate to the transport service. Admittedly, to design vehicles routes, it is desirable to take into account economic savings. Hence, the goal of designing is to select an optimization criterion (the most commonly used is the total distance traveled), which is to be minimized.

Problems aimed at designing optimal service vehicle routes can be divided into two typical tasks. Firstly, we require to search for the optimal trail (the optimal closed trail respectively) in the problem of *edge service of the graph*. This is discussed in detail in [4] and [5]. Secondly, the optimal route (the optimal cycle respectively) can be determined by solution of task focused on *vertex service of the graph*. Especially, this paper submits the second approach.

With regard to a practice, the frequent issue of collection distribution tasks are customers requirements located at the vertices of the transport network (graph respectively) and one of the theoretical methods to solve this problem can be so called *Traveling Salesman Problem* – a specific challenge is determining the minimum Hamiltonian circuit (MHC). Service vehicle route appropriated to MHC corresponds with such real situation where vehicle leaves the depot (initial vertex), serves each customer exactly once (all vertices of the graph) and returns to the depot.

Importantly, the essential part of each route designing is to respect real practical restrictions – the capacity of service vehicles as an example. The basic condition of finding MHC is following: the sum of customer's requirements must to be less than or equal to service vehicle capacity. Nevertheless, this condition may not be fulfilled

[1]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, marek.pomp.@vsb.cz
[2]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, petr.kozel@vsb.cz
[3]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, sarka.michalcova@vsb.cz
[4]Technical University of Ostrava/Faculty of Mining and Geology, Institute of Geoinformatics, 17. listopadu 15, Ostrava - Poruba, Czech republic, lucie.orlikova@vsb.cz

and so a large set of customers would not be fully satisfied immediately for the first time. As a consequence, it is necessary to design a set of circle roads. It is possible to use exact methods (for instance a mathematical programming) in case of small-scale tasks, the heuristic approach is beneficial to large-scale problems, [2]. Submitted contribution deals with the design of circle roads using one of the methods Cluster-First Route-Second in a real problem.

2 Formulation of problem

Let $N(V, E, l)$ be the graph with these attributes: $V = 1, ..., m$ is the set of vertices, $E = 1, ...n$ is the set of edges and l_{ij} represents edge evaluation ij, $i, j \in V$ as a road length (in kilometers). A depot with service vehicles r_k is located in the first vertex of the graph, where $k \in R$ and $r = 1, ..., p$ is a set of service vehicles with uniform capacity C. Customers z_j with requirements b_j are situated at vertices $j = 2, ..., m$, simultaneously with following:

$$\sum_{j=2}^{m} b_j \leq C \cdot p. \tag{1}$$

The goal of this task is designing individual circle routes for vehicles r_k to minimize the total distance traveled while serving customers z_j. The following three conditions must be met at the same time: every customer z_j is completely served by the only one visit of service vehicle r_k, the capacity of service vehicles C must not be exceeded and finally each of vehicles r_k must be used at most once.

So called Cluster-First Route-Second heuristic approach can be used to solve the problem mentioned earlier and it uses the following steps. At first, it is a clustering of the vertex set $S_1, ..., S_q$, where $S_1 \cup, ..., \cup S_q \cup \{d\} = V$. Furthermore, the following condition must be met: the sum of customer‘s requirements included to individual clusters does not exceed the capacity of service vehicle:

$$\sum_{j \in S_k} b_j \leq C, \text{ for } k = 1, ..., q. \tag{2}$$

Next, finding MHC is solved with some appropriate approach depending on task scale (exactly or heuristically).

3 The Sweep Algorithm

The Sweep Algorithm, which was introduced by Billy E. Gillet and Leland E. Miller in 1974, can be used for the vertex clustering, [3]. Here is the basic principal of this method. Every vertex i of the set $V = 1, ..., m$ is determined uniquely in a plane by a pair of numerical (Cartesian) coordinates, which are signed $[\bar{x}, \bar{y}]$. These coordinates $[\bar{x}, \bar{y}]$ must be transformed into coordinates $[x, y]$ with new origin of coordinate system $D[d_1, d_2]$ that corresponds to the depot. By using the following relationships, the transformation can be written as (3) and (4):

$$x = \bar{x} - d_1, \tag{3}$$

$$y = \bar{y} - d_2. \tag{4}$$

Thus, modified Cartesian coordinates $[x, y]$ are further converted to standard polar coordinates $[\varphi, r]$ by (5) and (6):

$$r = \sqrt{x^2 + y^2}, \tag{5}$$

$$\varphi = \arctan(y/x), \tag{6}$$

where r is the distance from origin to the point $[x, y]$ in general and φ is an angle measured counterclockwise from the positive x axis. Moreover, the conversion of φ for individual quadrants is following:

$$\varphi = \begin{cases} \arctan(y/x) & \text{for} \quad x > 0 \quad \wedge \quad y \geq 0\,, \\ \pi/2 & \text{for} \quad x = 0 \quad \wedge \quad y > 0\,, \\ \arctan(y/x) + \pi & \text{for} \quad x < 0\,, \\ 3\pi/2 & \text{for} \quad x = 0 \quad \wedge \quad y < 0\,, \\ \arctan(y/x) + 2\pi & \text{for} \quad x > 0 \quad \wedge \quad y < 0\,. \end{cases}$$

Furthermore, individual vertices must be sorted in ascending order (according to φ) with respect to the initial position from which clustering algorithm begins. The Sweep Algorithm for clustering does not have to start from the angle 0 in fact, so from the positive x axis, but formation of clusters can start at any angle $\alpha \in \langle 0, 2\pi)$. Clusters of vertices can be formed as follows:

- Service vehicle capacity C is specified and vertices are sorted to q clusters respecting restrictions (2).
- Number of clusters q is specified and vertices are evenly divided to these clusters regardless to fulfillment of the restrictive conditions (2).

3.1 Model examples

Cluster creation using the Sweep Algorithm can be illustrated on two trivial examples. There are 24 points in Cartesian coordinate system in the plane. A customer with the unit service requirement is located in each of these points. The goal is following:

- To create q clusters assuming that service vehicle capacity is $C = 6$ units fulfilling the restrictive conditions (2). The depot is situated outside the set of customers.
- To create 3 clusters of points in a way that numbers of customers assigned to individual clusters are uniformly distributed regardless to fulfillment of the restrictive conditions (2). The depot is situated inside the set of customers.

After transforming the Cartesian coordinates into the coordinate system with a new origin and subsequent transformation to polar coordinates, these polar coordinates are in ascending order according to φ and r. The clusters of points are subsequently created in prepared data.

Figure 1 shows the solution corresponding to both model examples. Firstly, the solution where 24 points are divided into 4 clusters, is on the left. Each cluster contains 6 points. The depot is marked with a square and initial angle was 0 while clustering. Secondly, the solution that corresponds to creating 3 uniform clusters with 8 points. Similarly, the depot is marked with a square and the initial angle was $\pi/3$ while cluster creating. Vertices, which belong to individual clusters, are marked by the same symbols.

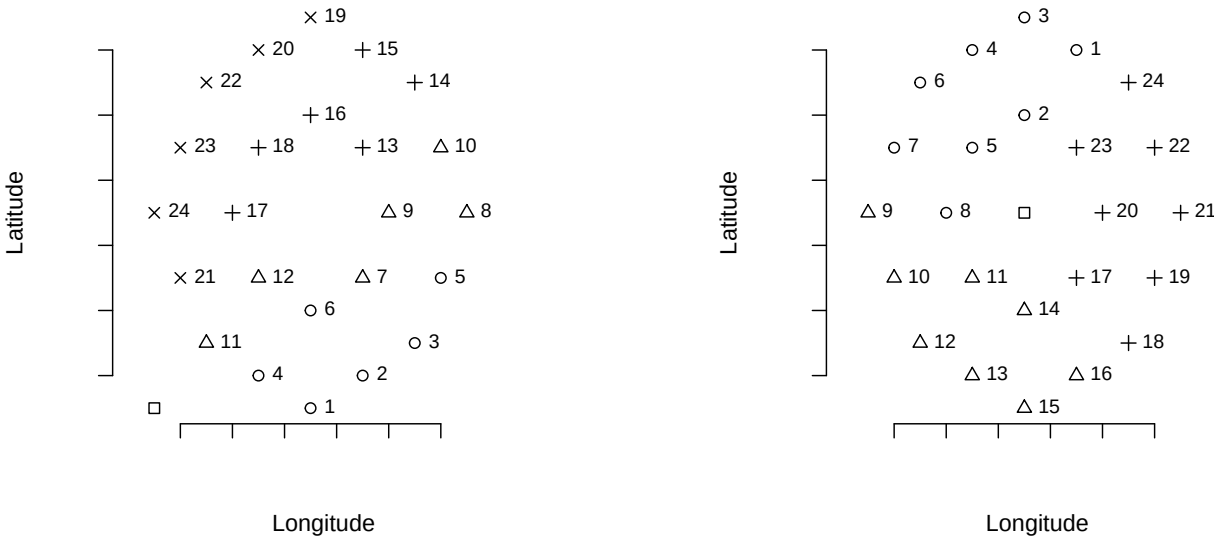


Figure 1 Solution of model examples

4 Numerical experiments

Numerical experiments using The Sweep Algorithm will be presented below. For computation were used the data of a private company in Moravian-Silesian Region dealing with municipal waste collection.

4.1 Description of real task

A private company deals with the municipal waste collection in a specific municipality and there are 1462 customers, which must be served. These customers are divided into 5 groups corresponding to specific collection interval. Intervals are as follows: *once a week* (1x7), *once every two weeks* (1x14), *once every three weeks* (1x21), *special service* (kx20) and *special service* (kx30). At the same time, each of customers is just in one group and moreover each of these 5 groups is divided to 3 subsets corresponding to such calendar days when the waste collection is realized (*Monday*, *Tuesday*, *Wednesday*). Again, each customer belongs to just one subset. This description is summarized in Table 1.

Day of collection	Monday	Tuesday	Wednesday	**TOTAL**
Collection interval	[-]	[-]	[-]	[-]
1x7	125	73	95	**293**
1x14	299	181	232	**712**
1x21	115	83	100	**298**
kx20	30	29	79	**138**
kx30	9	5	7	**21**

Table 1 Number of customers in individual subsets according to day of collection and collection interval

Subsets corresponding to both the collection interval and customer's needs are firmly determined. Nonetheless, the assignment of customers to the individual calendar days of collection is based only on the experience of dispatchers.

The goal of numerical experiments is to design an alternative location of customers to 3 subsets (3 calendar days of collection) in each of 5 groups corresponding to a collection interval. The capacity of service vehicles is quite sufficient for servicing all customers in appropriate calendar day of collection. Thus, one day corresponds to one circle route starting and ending at the depot. Furthermore, the important requirement of a company is a uniform location of numbers of customers assigned to individual calendar days.

4.2 Realization of experiments

In fact, a company providing the municipal waste collection has more data besides summarized data in Table 1: information about customers locations, customer's requirements and the depot location. Information about customer's location is available in form of postal address. Relevant postal addresses were geocoded using geoinformation tool *ArcGIS*, [1], [6] and each location has been assigned by GPS coordinations that can be considered simplistically as a customer position in the Cartesian coordinate system. And as mentioned earlier in 3, the Sweep Algorithm was applied to the input data. The implementation of the Sweep Algorithm was realized using computational tool *R*, [7]. Three uniform clusters S_1, S_2, S_3 of customers were designed for each of subsets (1x7), (1x14), (1x21), (kx20) and (kx30) using the Sweep Algorithm.

Then, the vertex corresponding to the depot $S_k \cup \{d\}$, where $k = 1, ..., 3$, was assigned to each cluster and also distance matrices c_{ij} corresponding to the travel network were determined using geoinformation tool *ArcGIS*. Furthermore, the task of finding MHC was solved heuristically in each of designed clusters using computational environment *Mathematica*, [8], [9]. The obtained results were compared with the current state and summarized in Tables 2 – 6.

Default state	Vertices	Length of circle road	Designe	Vertices	Length of circle road
	[-]	[km]		[-]	[km]
Monday	125	26.7	S_1	98	15.9
Tuesday	73	25.2	S_2	98	36.6
Wednesday	95	35.5	S_3	97	18.1
TOTAL	**293**	**87.4**	**TOTAL**	**293**	**70.6**

Table 2 Comparing the design with the current state for the collection interval(1x7)

Each of Tables 2 – 6 summarizes a comparison of the current state and a new design for the individual subset corresponding to intervals of collection (1x7), (1x14), (1x21), (kx20) and (kx30). Mentioned tables show information about numbers of served vertices, customers respectively, both in initial state and in the new proposal. In addition, each table gives an information about the length of circle road that must be realized to serve vertices (customers) - in kilometers. Admittedly, it is suitable to emphasize that while comparing the results in Tables 2 – 6 it is

Default state	Vertices [-]	Length of circle road [km]	Designe	Vertices [-]	Length of circle road [km]
Monday	299	49	S_1	238	19.4
Tuesday	181	37.8	S_2	237	42.6
Wednesday	232	53.3	S_3	237	26.8
TOTAL	**712**	**140.1**	**TOTAL**	**712**	**88.8**

Table 3 Comparing the design with the current state for the collection interval (1x14)

Default state	Vertices [-]	Length of circle road [km]	Design	Vertices [-]	Length of circle road [km]
Monday	115	37	S_1	100	16
Tuesday	83	30.3	S_2	99	35.8
Wednesday	100	47.6	S_3	99	27.2
TOTAL	**298**	**114.9**	**TOTAL**	**298**	**79**

Table 4 Comparing the design with the current state for the collection interval(1x21)

Default state	Vertices [-]	Length of circle road [km]	Design	Vertices [-]	Length of circle road [km]
Monday	30	18.1	S_1	46	23.6
Tuesday	29	19.9	S_2	46	21.6
Wednesday	79	35.7	S_3	46	18.2
TOTAL	**138**	**73.7**	**TOTAL**	**138**	**63.4**

Table 5 Comparing the design with the current state for the collection interval (kx20)

Default state	Vertices [-]	Length of circle road [km]	Design	Vertices [-]	Length of circle road [km]
Monday	9	9.2	S_1	7	6.5
Tuesday	5	6.7	S_2	7	10.1
Wednesday	7	5.9	S_3	7	6.5
TOTAL	**21**	**21.8**	**TOTAL**	**21**	**23.1**

Table 6 Comparing the design with the current state for the collection interval (kx30)

not possible to compare potential savings between individual days of collection and newly designed clusters. The main reason is that numbers of served vertices are not identical. Nevertheless, potential savings can be compared in the total sum. A summary of the achieved savings is provided in v Table 7.

Collection interval	Default state [km]	Design [km]	Total savings [km]
1x7	87.4	70.6	16.8
1x14	140.1	88.8	51.3
1x21	114.9	79	35.9
kx20	73.7	63.4	10.3
kx30	21.8	23.1	-1.3
TOTAL			**113**

Table 7 Summary of total savings

Table 7 shows that thanks to the newly designed clusters, the service of customers with savings (in kilometers) can be realized in all four subsets corresponding to intervals of collection (1x7), (1x14), (1x21) and (kx20). In contrast, the new proposal has caused a deterioration of the initial state in the last subset (kx30), but this deterioration is not significant (**-1.3** km). As a result, the savings in the new proposal were **113** km.

5 Conclusion

The presented contribution has been dealing with the design of circle roads using the Cluster-First Route-Second method in conditions of real problem. In this paper, the challenge was aimed at using the Sweep Algorithm to decomposing a wider set of vertices. This method was described in detail in chapter 3. Moreover, the task of finding the minimum Hamiltonian circuit (MHC) was subsequently solved in individual clusters created using the Sweep Algorithm. This previously mentioned approach was applied to the specific real task and all results of experiments were summarized in 4. The results of experiments show that service of customers with savings **113** km can be realized thanks to newly designed clusters.

Acknowledgements

This paper was supported within Op. Prog. Education for Competitiveness Project No. CZ.1.07/2.3.00/20.0296 and programme for the support of applied research and experimental development, Technology Agency of the Czech Republic (EPSILON): Effective approaches to economical and adaptable systems of maintenance and operation of transport networks TH02010930.

References

[1] Allen, D., W.: *GIS Tutorial 2: Spatial Analysis Workbook*. Esri Press, New York, 2013.

[2] Bodin, L., Golden, B., et al.: *Routing and Scheduling of Vehicles and Crews*. Pergamon Press, Oxford, 1983.

[3] Janáček, J.: *Optimalizace na dopravních sítích*. Žilinská univerzita, Žilina, 2002.

[4] Kozel, P., Michalcová, Š.: The use of linear programming in solving service tasks of mixed transport networks. In: *Proceedings of 11th International Conference Strategic Management and its Support by Information Systems 2015*, Uherské Hradiště, 2015, 274–279.

[5] Kozel, P., Michalcová, Š.: The use of linear programming to solve routing task in practice. In: *Proceedings of 33th International Conference Mathematical Methods in Economics 2015*, Cheb, 2015, 395–400.

[6] Ormsby, T., Napoleon, E. et al.: *Getting to Know ArcGIS Desktop*. Esri Press, New York, 2010.

[7] *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, 2016.

[8] Skiena, S., and Pemmaraju, S.: *Computational Discrete Mathematics*. Cambridge University Press, New York, 2003.

[9] Wolfram, S. *The Mathematica Book*. Cambridge University Press, New York, 1999.

Optimization of the tasks and virtual machines allocation problem

Daniela Ponce[1], Vladimír Soběslav[2]

Abstract. The tasks and virtual machines allocation are one of the most important problems in efficient use of IT infrastructure including the physical, virtualization or computing technology. Depending on the exact setup of the computing scenario, the optimization problem can be stated in different ways. Most optimization works concern single data centers, multiple infrastructure-as-a-service (IAAS) systems, virtual machine - physical machine assignment problem and task allocation.
Much work at the above problems has been done and interesting results have been found. Yet, several aspects were neglected and need to be addressed in more details. Traditional distribution of task or virtual machine allocation is based on simple load balancing techniques, e.g. round robin, in coming data stream selection, etc. This kind of distribution does not reflect the real utilization of resource or problem such as communication between virtual machines, virtual machine co-location interference, composition of power consumption optimization at multiple levels of components, or physical machine with multi-core CPU. The paper is focused on implementation of online bin packing algorithm for workload optimization based on computing task categorization.

Keywords: cloud computing, resource allocation, optimization.

JEL classification: C44
AMS classification: 90C27

1 Introduction

Currently we are witnessing a significant increase of information technologies utilization, consequently the number of tasks that are processed or automated by the computer is gradually rising. This trend reflects the increase use of virtualization platforms such as Microsoft Hyper-V, VMWare, XEN, KVM and also is related with the development of cloud computing technologies, such as IaaS - Infrastructure as a Service, PaaS - Platform as a Service or SaaS - Software as a Service.

"One of the challenges posed by cloud applications is Quality-of-Service (QoS) management, which is the problem of allocating resources to the application to guarantee a service level along dimensions such as performance, availability and reliability" [1]. Regardless of the technologies used, one of the critical factors for efficient use of physical, virtualized or cloud computing resources is the allocation of tasks between individual computational nodes [20]. This is because the standard methods of task distribution between servers are unable to distinguish between demanding and simple tasks. In the real-world scenario, application and tasks which are normally allocated on compute nodes are not identical and the workload is changing dynamically over time. Workload has usually nonlinear characteristics. Individual tasks vary in complexity and have a different impact on hardware components of the computational node, especially differences are in CPU usage, memory allocation, I/O subsystem, graphics cards, networking utilization and much more. Following the previous mentioned categorization, we can name Database intensive task, File sharing task, Application server tasks etc. According to [18], we can distinguish the categories below:

- CPU intensive tasks - relatively small amount of data on compute node, complex algorithms, corresponding memory and I/O allocation.
- in memory intensive tasks - high memory allocation on compute node, CPU intensive, corresponding I/O allocation.
- I/O intensive tasks - huge amount of stored data, storage or database intensive tasks.

According to the different computing types in cloud computing mentioned above, the cloud computing infrastructure (IaaS) should be optimized for specific systems workloads. Traditional or even the cloud computing infrastructure is composed of many elements intended to cooperate together. Another, yet important factor is the

[1]University of Hradec Králové, Dept. of Information Technologies, Rokitanského 62, Hradec Králové, Czech Republic, daniela.ponce@uhk.cz
[2]University of Hradec Králové, Dept. of Information Technologies, Rokitanského 62, Hradec Králové, Czech Republic, vladimir.sobeslav@uhk.cz

electricity consumption, for example DC electricity consumption in the USA alone will increase to 140 billion kWh per year by 2020, this means 13 billion USD annually in electricity bills and nearly 100 million tons of CO2 per year [19]. Power consumption can be significantly reduced by the efficient workload allocation and reduction to the minimum number of necessary compute nodes. Especially cloud computing infrastructure is capable of virtual machine live migration or increase. Traditional workload allocation/task distribution is handled by load balancing techniques. The biggest problem of traditional load balancing algorithms is the distribution logic which is usually not related with the utilization of hardware resources on compute nodes. The distribution algorithms dont reflect the differences between specific tasks and their intensity.

The aim of this paper is to propose load balancing solution which is based on on-line bin packing algorithm and task categorization. Following chapters cover survey of load balancing methods, technologies and finally the bin packing algorithm which was tested on selected data set from Oracle Real Application cluster.

2 Survey of allocation problem

Although we mention virtual machine allocation and tasks allocation jointly, these two problems are slightly different. While in virtual machine allocation we aim at maximal number of virtual machines at one physical machine and we even try to reduce the number of physical machines in use by virtual machines consolidation, it is not possible to reallocate tasks being performed and we have also to avoid overloading of compute nodes since compute node recovery results in temporary lessening of the available computing capacity.

2.1 Virtual machines allocation

Basically, the virtual machine allocation problem is formulated as a problem of allocation of virtual machines to physical machines in such a way that aims at selected objectives. There exist many variants of virtual machine allocation problem. The majority of them deals with two different problems [15], as seen from the point of view (and interest) of the cloud service provider: (1) the cloud service provider owns one data center with physical machines (single data center case), (2) the cloud service provider owns no physical machines and he acts as service broker (multi-Infrastructure-as-a-Service case). Few other works address the problem of cloud provider possessing more than one data center or the problem of cloud provider offering his own and external data centers, too.

Allocation of virtual machines is multi-objective problem. Some of the objectives are independent, others not, and some of them are even conflicting. It is common practice in definition of optimization problem to focus at one objective or to consider weighted sum of several objectives. As formulated in [15], objectives put in existing works can be grouped in the following manner:

- cost or income related (e.g. [3], [24]),
- performance related (e.g. [17], [23]),
- energy related (e.g. [2], [25], [26]), or
- technical parameters related (e.g. [13], [21], [22]).

In our model we focus at effective utilization of physical resources and load balance between physical machines in data centers which leads to reduction of bandwidth required by virtual machine.

Problem solution approaches include exact methods (exact algorithms, linear programming, binary integer programming, off-shelf solvers). In the case of single data center the virtual machine allocation problem can be viewed as bin-packing problem. Since bin-packing is NP-complete problem [8], metaheuristics have been also developed, e.g. simulated annealing, genetic algorithms, ant colony optimization [9]. For the multi-IaaS version of cloud computing, a plenty of heuristics has been designed: list scheduling, greedy provisioning and allocation policies, greedy scheduling and clustering algorithms, variations of algorithms for knapsack problem, particle swarm optimisation [15].

2.2 Tasks allocation

In paper [4], the problem of dynamic resource allocation to clustered web applications was studied. The application server middleware was added the ability to automatically decide the size of application clusters and their placement on physical machines. The approach considered the satisfaction of each application with a particular resource allocation and attempts to at least equally satisfy all applications. The satisfaction is modelled using utility functions, mapping CPU resource allocation to the performance of an application relative to its objective. The demonstrated online placement technique aimed at equalizing the utility value across all applications while also satisfying operational constraints, preventing the over-allocation of memory, and minimizing the number of placement changes.

In the work presented in [5], a time-domain description of a generalized processor sharing (GPS) server was used to model the server resources with the purpose to capture the transient behavior of the application workloads.

In this model, the application resource requirements were mapped to their dynamically changing workload characteristics. Online monitoring and prediction is used to update the parameters of this model. The expected workload parameters are estimated by time series analysis of the measured system metrics. The allocation of server resources is formulated as a constrained non-linear optimization task.

The authors of [10] propose an algorithm that aims at minimum cost and maximum flow for resources placement in clouds which has to deal with dynamic workloads and flows variations. The allocation problem for cloud resources is modeled as a directed graph. A finite number of resource types is supposed.

In paper [27], the problem of jointly allocating compute (i.e., processor and memory) resources and network resources in a large-scale cloud environment is considered. The problem of optimally allocating compute and networking resources to VDCs for four management objectives is formulated (balanced load, energy efficiency, fair allocation, service differentiation). An architecture for resource allocation under management objectives is presented consisting of a set of cooperating controllers, each solving a problem related to an objective.

3 Task allocation as bin packing problem

We focus our attention on compute nodes. They are characterized by computing capacity and services offered. The computing capacity can be measured along one or several dimensions, e.g. CPU capacity, memory size. The services offered by compute nodes and required by tasks are for example database operations, file handling operations, computing intensive operations.

The compute nodes perform tasks which they receive from load balancer. Once the task is sent by load balancer to a compute node, it can not by redirected to other compute node. If the compute node is completely busy, the task waits in the compute nodes buffer till its execution. The load balancer has to take decision with little information only since nor the computing capacity required by the current task is known before the execution of that task, nor it has real-time information about the free computing capacity of the compute nodes.

The objective of the load balancer is to allocate the tasks to compute nodes in such a manner that: the compute nodes are loaded with tasks evenly the compute nodes execute as much tasks as possible and, at the same time, they get overloaded as little as possible the tasks are performed as soon as possible the tasks are not lost because of compute node overloading

3.1 Problem definition

As we have mentioned in the introduction, the real-world task distribution and allocation does not reflect the differences between specifics of computing task itself and the workload characteristics. The standard methods for load balancing do not distribute computational tasks according to their complexity, but usually use a simple mechanisms such as Round Robin. For this reason, there must be a sufficient power reserve on the servers due to the unevenness of the specific tasks and their workload characteristics. Furthermore, load balancing algorithms do not have direct feedback from compute nodes. According to the level of intelligence we can categorize the load balancing method into the following categories:

- Round robin – It is a simple method for load balancing task between compute nodes. It provides simple fault tolerance. The topology is set from two or more compute nodes that work within one domain. These compute nodes are configured to process the tasks/requests. When the the load balancer receives the task/request, it distributes the tasks rotationally between computing nodes based on IP addresses or a similar identifier.
- Least Connection – Round Robin does not take into account the current compute node workload when distributing the requests. This method takes into account the current compute node workload. Requests come to the compute node that currently serves the least active sessions. This method allows to count the number of transactions running on the nodes, but their difficulty and workload characteristics are not considered.
- Adaptive agent-based load balancing – Each compute node in the pool has an agent that reports its current load to the load balancer. This real-time information is used to decide which compute node is best suited for processing the request. This method is dependent on the agent's intelligence and its ability to predict load, or to categorize single-task tasks.
- Chained Failover – In this method, a predetermined compute node order is configured in the chain/string. All requests are sent to the first chain compute node. If it can not accept any additional requests, the next compute node in the chain will handle the remaining requests.
- Delay Response Time – This method is based on the compute node response check and determines repeatedly which compute node responds most quickly within the given time. The next compute node request distribution is done after the load balancer evaluates the current compute nodes workload and consequently distributes the tasks according to the current workload.
- Source IP address – This kind of source load balancing is using a hash algorithm that takes the source and destination IP address of the client and compute node and combines them to generate a unique hash key.

This key is used to allocate the client to a specific compute node. This key can be restored if the session is interrupted. Source load balancing can ensure that the client request is routed to the same compute node which was used earlier. This is useful if client should establish session with a particular compute node.

If we could categorize the tasks in some way based on their real workload characteristics, we would be able to better decompose performance requirements and get closer to the optimal utilization of computational nodes. This approach would save money on purchasing hardware, software, technological support, energy, etc.

Based on the work [6], we can formulate the problem of task allocation as bin packing problem where bins are compute nodes and items to be packed are tasks. The capacity of bins and the size of items is measured in CPU performance and memory size (and possibly other characteristics). The item can be packed into a bin, if the size of the item is at most the remaining capacity of the bin in all dimensions (CPU performance, memory, etc.). We assume a **fixed** number of bins, therefore we aim to pack **max number of items**. Tasks are coming in an infinite stream, so we have **online** variant of bin packing. Since the load balancer locates a task without knowing its exact duration and load required from compute node, we do not have (exact) information regarding the size of items not yet packed and thus the problem is **stochastic**. Compute nodes differ in capacity, so the we have **variable-sized** bins. After their execution, tasks no longer occupy capacity of compute nodes, which means that we have **dynamic** bin packing where items can be deleted from bins after some time interval. The problem is also **class constrained**, since tasks can be sorted according to the type of computing service they require.

3.2 Problem solution

The bin packing problem is NP-hard (see, e.g., [8]). Therefore many approximation algorithms based on heuristics were developed, which do not guarantee an optimal solution for every instance, but attempt to find a near-optimal solution within polynomial time [6]. The classical heuristics include Next-Fit, Worst-Fit, First-Fit and Best-Fit [6]. They all compare the size of the current item to be packed with the capacity total or left in the current bin or in all bins and they also consider the indexing order of the bins. As they make use of information which is not available to the load balancer, they are not suitable for our problem. Another well-known heuristic is Harmonic-Fit [12]. The algorithm divides the size of items into k non-overlapping size intervals such that each size falls in exactly one size range. Items of size falling into the same range are assigned to the same bin and the bin is packed exclusively with the items assigned to it using the Next-Fit heuristic (which starts with an empty bin and packs items into the bin till the item to be packed fits in, otherwise it shifts to an empty bin) so at most k bins are being packed at the same time.

We propose three heuristics that can be used in approximation algorithm for the formulated bin packing problem: Stochastic-Harmonic, Multiple-Lanes and Free-Capacity variations.

Stochastic-Harmonic. Tasks are sorted according to the their expected computing capacity needs which can be detected based on their class membership. Since there are typically more task classes than compute nodes, task classes are grouped based on their frequency in such a way that task classes get evenly distributed among compute nodes with respect to the count of tasks in each task class. Similarly, the distribution can be done according to the expected workload relative to the compute node capacity.

Multiple-Lanes. Tasks are sorted into bins based on their expected duration. The tasks belonging to the class with the largest expected duration are distributed to a separate compute node; all other tasks are evenly distributed regardless their class.

Free-Capacity variations. As previous two heuristics, augmented with the consideration of current workload of the compute nodes.

3.3 Example of problem instance

We have collected workload data from an existing e-commerce company. The workload traces from one day were selected as a representation of typical workload distribution. The workload trace was categorized in to four types of tasks and the frequency of computational demands of selected categories was measured. The workload trace was collected on Oracle 11GR2 RAC (Real Application Cluster) which consists of two compute nodes. This database system servers as the main enterprise resource planning system. Figure 1 presents an example of Top database actions and activities. This categorization is based on the measurement and frequency of SQL statements (such as select, insert, update etc.) which represent the specific tasks/actions.

The table 1 contains information on task classes. For each task class, typical duration, frequency of tasks and computing capacity needs are given. Furthermore, division of computing capacity needs into low computing capacity (LCC) required tasks percentage and high computing capacity (HCC) required tasks percentage is given. Low computing capacity required by a task denotes less than 25% of total computing capacity available at compute node, while high computing capacity required by a task denotes more than 25% of total computing capacity

available at compute node.

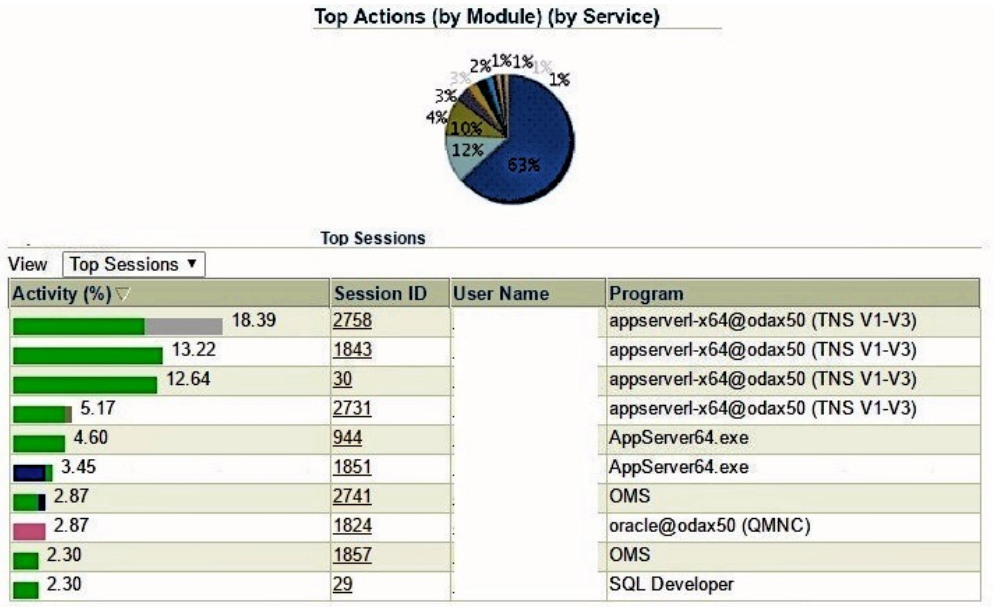


Figure 1 Top database actions and activities

class	duration (s)	frequency	LCC	HCC
1	0-2	63%	90%	10%
2	2-5	26%	80%	20%
3	5-15	10%	75%	25%
4	15-60	1%	60%	40%

Table 1 Tasks classification

4 Conclusion

We have formulated the task allocation problem as a dynamic, stochastic online bin packing problem. We have proposed several heuristics as a base of approximative algorithm. These heuristics do not require exact knowledge on individual tasks and they are based solely on statistical analysis of historical data regarding task distribution, duration, frequency and required source amount. As the next step, we will test the heuristics within some appropriate testing framework, i.e. using actual historical data, or benchmark data sets, eventually in real environment.

Based on the analysis of more than 250 research publications dealing with resource management in cloud environments, the authors of [11] identified several challenges for future investigation including *engineering scalable resource management systems*. Effective resource management remains in the focus of the research in the upcoming years, too.

Acknowledgements

The support of the specific research project SPEV named Computer networks for cloud and distributed computing at FIM UHK is gratefully acknowledged.

References

[1] Ardagna, D., Casale, G., Ciavotta, M. et al.: Quality-of-service in cloud computing: modeling techniques and their application. *Internet Serv Appl* **5**: *11* (2014). doi:10.1186/s13174-014-0011-3

[2] Beloglazov, A., Abawajy, J.,Buyya, R.: Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future Generation Computer Systems* **28** (2012), 755–768.

[3] Van den Bossche, R., Vanmechelen, K., Broeckhove, J.: IaaS reserved contract procurement optimisation with load prediction, *Future Generation Computer Systems* **53**, December (2015), 13–24.

[4] Carrera, D., Steinder, M., Whalley, I., Torres, J., Ayguade, E.: Utility-based placement of dynamic web applications with fairness goals. In *Proc. 11th IEEE/IFIP Network 38 Operations and Management Symposium (NOMS 2008)*, (2008), IEEE, 9-16.

[5] Chandra, A., Gong, W., Shenoy, P.: Dynamic resource allocation for shared data centers using online measurements. In *K. Jeffay, I. Stoica, K. Wehrle (eds.) Proc. 2003 International Workshop on Quality of Service (IWQoS 2003)*, LNCS **2707** (2003), 381-398.

[6] Coffman, E. G., Csirik, J., Galambos, G., Martello, S., Vigo, D.: Bin Packing Approximation Algorithms: Survey and Classification. In *Pardalos, P. M., et al. (eds.): Handbook of Combinatorial Optimization, Springer, New York 2013*, 455–531.

[7] Ficco, M., Di Martino, B., Pietrantuono, R., Russo, S.: Optimized task allocation on private cloud for hybrid simulation of large-scale critical systems. *Future Generation Computer Systems* (2016), doi.org/10.1016/j.future.2016.01.022.

[8] Garey, M. R., Johnson, D. S.: *Computers and Intractability: A Guide to the Theory of NPCompleteness.* W.H. Freeman, New York, 1979.

[9] Guzek, M., Bouvry, P., Talbi, E. G.: A survey of evolutionary computation for resource management of processing in cloud computing [review article]. *IEEE Computational Intelligence Magazine* **10** (*2*), (2015) 53–67.

[10] Hadji, M., Zeghlache, D.: Minimum cost maximum flow algorithm for dynamic resource allocation in clouds. In *Proceedings of the 2012 IEEE 5th International Conference on Cloud Computing, CLOUD 12, Honolulu, HI, USA* (2012), 876-882.

[11] Jennings, B., Stadler, R.: Resource Management in Clouds: Survey and Research Challenges. *Netw Syst Manage* **23** *567* (2015). doi:10.1007/s10922-014-9307-7

[12] Lee, C. C., Lee, D. T.: A simple on-line bin-packing algorithm. *Journal of the ACM* **32**(*3*), (1985), 562–572.

[13] Levitin, G., Xing, L., Dai, Y.: Optimal data-partitioning in cloud computing system with random server assignment.*Future Generation Computer Systems* **70**, May (2017), 17–25.

[14] Malawski, M., Figiela, K., Nabrzyski, J.: Cost minimization for computational applications on hybrid cloud infrastructures. *Future Generation Computer Systems* **29**, 7 (2013), 1786–1794.

[15] Mann, Z. A.: Allocation of virtual machines in cloud data centers: a survey of problem models and optimization algorithms. *ACM Computing Surveys* **48**, 1 (2015), 11:1–11:34.

[16] Mann, Z. A.: A taxonomy for the virtual machine allocation problem. *International Journal of Mathematical Models and Methods in Applied Sciences* **9** (2015), 269–276.

[17] Mann, Z. A.: Rigorous results on the effectiveness of some heuristics for the consolidation of virtual machines in a cloud data center. *Future Generation Computer Systems* **51** (2015), 1–6.

[18] Marešová, P., Soběslav, V., Klímová, B.: Methodological Approach to Efficient Cloud Computing Migration. In *Studies in computational intelligence. 2015, 598 (2015), 199-208, 191-201.* ISSN 1860-949X.

[19] Martens, B., Walterbush, M., Teuteberg, F.: Costing of Cloud Computing Services: A Total Cost of Ownership Approach. In *Proceedings of the 45th Hawaii International Conference on System Sciences, Maui, Hawaii, USA 2012, 15631572.*

[20] Pavlík, J., Soběslav, V., Komárek, A.: Measurement Of Cloud Computing Services Availability. In *Nature of computation and communication (ICTCC 2014), Springer, New York 2015, 191-201.* ISBN 978-3-319-15392-6.

[21] Sindelar, M., Sitaraman, R. K., Shenoy, P.: Sharing-aware algorithms for virtual machine colocation. In *Proceedings of the twenty-third annual ACM symposium on Parallelism in algorithms and architectures*. ACM (2011), 367–378.

[22] Slegers, J., Mitrani, I., Thomas, N.: Evaluating the optimal server allocation policy for clusters with in/off sources. *Performance evaluation* **66**, 8, August (2009), 453–467.

[23] Srikantaiah, S., Kansal, A., Zhao, F.: Energy aware consolidation for cloud computing. *Cluster Computing* **12** (2009), 1–15.

[24] Sun, Y., White, J., Eade, S., Schmidt, D. C.: ROAR: a QoS-oriented modelling framework for automated cloud resource allocation and optimization. *Journal of Systems and Software* **116** (2016), 146–161.

[25] Vakilinia, S., Heidarpour, B., Cheriet, M.: Energy efficient resource allocation in cloud computing environments, *IEEEAccess* **4** (2016), 8544–8557.

[26] Verma, A., Ahuja, P., Neogi, A.: pMapper: power and migration cost aware application placement in virtualized systems. In *Middleware 2008*, 2008, 243–264.

[27] Wuhib, F., Yanggratoke, R., Stadler, R.: Allocating Compute and Network Resources Under Management Objectives in Large-Scale Cloud. *J Netw Syst Manage* **23** (2015), 111-136.

The spatial weights matrices and their influence on the quality of spatial models of employment

Ewa Pośpiech[1], Adrianna Mastalerz-Kodzis [2]

Abstract. In the article the problem of employment modelling is analysed. The analyses have been carried out taking into account the spatial factor. To take into account of this factor means to use the spatial weights matrices, that define the spatial structure of the neighbourhood and that are an essential part of the research. The studies included regions of selected countries of Central Europe. The role of the explained variables played the characteristics describing the number of employees in total and the number of employed men and women, while the explanatory variables were selected demographic characteristics and GDP per capita. The studies included two basic spatial models: spatial lag model and spatial error model – in each case, a better tool of analysis turned out to be the spatial error model. In the analyses different spatial weights matrices were considered and the best of the variants were indicated.

Keywords: spatial autocorrelation, spatial modelling, spatial weights matrix, employment.

JEL Classification: C33, C51, C52
AMS Classification: 91B40, 91B72

1 Introduction

In the case of issues that are embedded in a particular area (such as: demographic processes, phenomena occurring in the real estate market, the labour market or economic development[3]) the spatial factor may be important element of the analysis[4]. It can significantly affect the quality of the studied models and improve their properties. In situation, when the significance of this factor is identified, the appropriate (spatial) methods should be taken into account in the analysis. The use of these methods means to take into account of the spatial weights matrices. These matrices define the structure of spatial neighbourhood and they are an important part of the analyses. The spatial weights matrices can be defined in different ways, and the choice of one of them in a significant way can affect the results.

One of the essential markets in each countries' economy is the labour market. Characteristics like unemployment rate or employment level (in total as well as according to gender) are conscientiously observed and constantly analysed – they are essential determinants of policy of the regions and countries. Monitoring of this phenomena, identification of new factors shaping the level of the characteristics and the ability to manage these factors facilitate efficient socio-economic policy – this allows to prevent undesirable phenomena and is conductive to maintaining the positive trend. Analyses of the phenomena that shape the labour market (using modern tools), are indispensable.

The purpose of the paper is to identify the significance of spatial factor in the description of the level of employment of selected regions of Central Europe and to examine the influence of the choice of the spatial weights matrix on the quality of the models of employment. The article is divided into two main parts: theoretical one and empirical one. In the theoretical one the stages of identification of the spatial factor and considered weights matrices types were discussed whilst in the empirical one the received results and conclusions were presented.

[1] University of Economics in Katowice, Department of Mathematics, 1 Maja 50, 40-287 Katowice, ewa.pospiech@ue.katowice.pl
[2] University of Economics in Katowice, Department of Mathematics, 1 Maja 50, 40-287 Katowice, adrianna.mastalerz-kodzis@ue.katowice.pl
[3] A discussion of these issues can be found e.g. in [5], [6], [7], [8], [9].
[4] The significance of this factor is indicated by Tobler's law – the law says that neighbouring locations are more similar concerning the studied feature than locations further apart from each other [11].

2 Spatial modelling

2.1 Identification of spatial models

Spatial modelling requires appropriate methods of estimation. An important role in these methods play matrices of neighbourhood – the matrices influence on the explained variable in a different way (depending on which spatial model is estimated). We can enumerate two basic spatial models: spatial lag model – SLM, and spatial error model – SEM. We can also mention spatial cross-regressive models, mixed variants and others [10]. The models used in the article are the basic ones and these are described below.

In the first one – SLM, the main component is the spatially lagged explained variable **Wy**. This model can be written as follows:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \tag{1}$$

where: ρ – spatial autocorrelation coefficient,
$\mathbf{W}$ – spatial weights matrix,
$\boldsymbol{\beta}$ – the vector of model coefficients,
$\mathbf{X}$ – the matrix of the explanatory variables.

In the SEM model the important component is the spatially lagged error $\mathbf{W}\boldsymbol{\xi}$ (a spatial autocorrelation of the residuals is assumed in the model). The form of the model is expressed as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi} \tag{2}$$

$$\boldsymbol{\xi} = \lambda\mathbf{W}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \tag{3}$$

where: λ – spatial autocorrelation coefficient, other symbols as in the above.

The process of identification of the correctness of using spatial modelling takes part in stages (according to [1], [2], [3], [4], [10]). As the first one we can enumerate the linear econometric model estimation using the ordinary least square method (OLS). The obtained model is then analysed in terms of the presence of spatial autocorrelation of the model residuals. This can be accomplished using the Moran's *I* test for residuals. The statistic for the test is defined as follows:

$$I = \frac{n}{S_0} \cdot \frac{\mathbf{u}^{\mathbf{T}}\mathbf{W}\mathbf{u}}{\mathbf{u}^{\mathbf{T}}\mathbf{u}}, \tag{4}$$

where: n – the number of included regions,
$\mathbf{u}$ – n-dimensional column vector of the model residuals,
S_0 – the sum-total of weights matrix, $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$,
$\mathbf{W}$ – spatial weights matrix.

With the use of Moran's *I* test the following hypotheses are verified – H_0: no spatial autocorrelation of the model residuals; H_1: the presence of spatial autocorrelation of the model residuals. However, the occurrence of spatial autocorrelation of the model residuals does not provide a guarantee for using spatial methods in modelling considered phenomenon – it can mean, for example, incorrect specification of the model. Then, the analysis should be supported by a graphical representation illustrating the spatial distribution of the residuals. Moreover, the join-count test should be used in the analyses. The test enables the assessment of random (or non-random) distribution of the residuals – the non-random distribution suggests that there are other factors affecting considered variables.

The next stage in the identification of the spatial model is carrying out of Lagrange multiplier diagnostic tests (*LM* tests). These tests that are based on OLS residuals enable the indication of the proper spatial model. We can distinguish the ordinary tests and robust ones. In the ordinary tests the null hypothesis (for the spatial error model) proposes that there is no spatial autocorrelation (H_0: $\lambda = 0$), whilst the alternative hypothesis assumes that there is spatial autocorrelation (H_1: $\lambda \neq 0$). The statistic LM_{ERROR} with asymptotic distribution $\chi^2(1)$ is defined as follows:

$$LM_{ERROR} = \frac{1}{T_1}\left(\frac{\mathbf{u}^{T}\mathbf{W}\mathbf{u}}{\hat{\sigma}^2}\right)^2 \tag{5}$$

where: $\hat{\sigma}$ – estimated standard error,
T_1 – parameter expressed in the formula: $T_1 = \mathrm{tr}[(\mathbf{W}^{\mathrm{T}} + \mathbf{W})\mathbf{W}]$, other symbols as in the above.

In the case of spatial lag model the LM_{LAG} statistic's value with asymptotic distribution $\chi^2(1)$ is calculated using the following formula:

$$LM_{LAG} = \frac{1}{T_2}\left(\frac{\mathbf{u}^T\mathbf{W}\mathbf{y}}{\hat{\sigma}^2}\right)^2 \tag{6}$$

where: T_2 – parameter expressed in the formula:

$$T_2 = T_1 + \frac{(\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}})^T\mathbf{M}(\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}})}{\hat{\sigma}^2} \tag{7}$$

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \tag{8}$$

and

$\hat{\beta}$ – the estimated vector of coefficients, other symbols as in the above.

The statistic expressed in the formula (6) enables the verification of the hypothesis on the significance of ρ coefficient (H_0: ρ = 0, H_1: ρ ≠ 0).

The selection of the model (according to hybrid strategy) is based on the higher and statistically significant value of *LM* statistic: if $LM_{ERROR} > LM_{LAG}$, the spatial error model should be appointed; otherwise (when $LM_{LAG} > LM_{ERROR}$), the spatial lag model should be estimated. When the information based on the ordinary tests is not unanimous, the robust tests should be applied.[5]

An important element in the procedure of the identifying of the spatial model is carrying out of additional diagnostic tests. We should indicate such tests as likelihood ratio test (*LR* test) and Wald's test. Comparing values of these tests statistics with the values of *LM* statistics we can determine whether the model specification is appropriate – if we have the following inequality: Wald(λ) ≥ LR_{ERROR} ≥ LM_{ERROR}, it means that the spatial error model is well specified; on the other hand, if the following dependence is satisfied: Wald(ρ) ≥ LR_{LAG} ≥ LM_{LAG}, the spatial lag model is properly specified.

2.2 Spatial weights matrices

As already mentioned above, an important role in the spatial methods play the weights matrices. They define the structure of spatial neighbourhood, they measure the spatial relationship as well as represent the power of the potential interactions between locations. The matrices are built to specify the spatial dependence, they are some kind of restriction imposed on spatial structure of neighbourhood. We can distinguish many types of weights matrices. The most common way to present the structure of the neighbourhood is adopting as a criterion of nearness the contiguity criterion. We can enumerate several types of matrices connected with this criterion. The basic matrix is basic binary coding one $\mathbf{B} = [w_{ij}]$, $i = 1, \ldots, n, j = 1, \ldots, n$, of the following elements [4]:

$$w_{ij} = \begin{cases} 1, & \text{when objects } i \text{ and } j \text{ share a border} \\ 0, & \text{when objects } i \text{ and } j \text{ do not share a border} \\ 0, & \text{for diagonal elements} \end{cases}. \tag{9}$$

Other types are: the row standardised first-order contiguity matrix **W** – is the most popular weights matrix type; the globally standardised matrix **C**; **U** matrix – **C** matrix divided by the numbers of neighbours; the variance stabilizing matrix **S** (this matrix is used in percentage analyses – when the number of people in studied regions is different). Another type of matrix is the *k*-nearest neighbours matrix – the nearest neighbour is the one whose centre is located close to the centre of the studied area (on the basis of the Euclidean distance). The next type of spatial weights matrix can be a matrix of neighbours within *d* km – the neighbour is the area that the centre is not more than *d* km from the centre of studied area. We can also talk about the weights matrix based on the criterion of the inverse of the distance, etc. Among non-standard matrices we can distinguish, for example, the Cliff and Ord matrix, social distance matrix or economic distance one.

3 Empirical analysis

3.1 Subject and object of the research

The subject of the research were the regions of six selected European countries, specified on the base of EU classification, NUTS2 system. The following countries were taken into account: Austria, the Czech Republic, Germany, Hungary, Poland, Slovakia. The empirical analysis was conducted for 2011 and 2015 and involved the following variables:

[5] Test statistics for robust tests could be found e.g. in [10].

- Employment of the age 15-64 (thousands), total – ET;
- Employment of the age 15-64 (thousands), males – EM;
- Employment of the age 15-64 (thousands), females – EF;
- Regional gross domestic product (PPS per inhabitant) – GDP;
- Population density (persons per square kilometre) – PD;
- Population, age from 15 to 64 years (thousands), total – PT;
- Population, age from 15 to 64 years (thousands), males – PM;
- Population, age from 15 to 64 years (thousands), females – PF.

The data were sourced from the Eurostat base, while the calculations were developed using the MS Excel and R CRAN package.

3.2 Spatial models of employment

In the first stage of the research linear econometric models were estimated. The explained variables were the number of employed persons: the total number and according to gender. Using ordinary least square method, three models for each of the selected years were constructed. These models were analysed in respect of spatial properties – it was analysed whether using spatial modelling is justified. The estimated models parameters, using OLS method, are presented in table 1.

Year	Explained variable	Explanatory variables					Intercept	R^2
		GDP	PD	PT	PM	PF		
2011	ET	7.76E-03	-2.91E-02	6.65E-01	×	×	-1.77E+02	0.979
	EM	3.77E-03	-1.98E-02	×	7.26E-01	×	-8.67E+01	0.987
	EF	4.01E-03	-9.32E-03	×	×	6.05E-01	-9.17E+01	0.967
2015	ET	5.80E-03	-2.92E-02	6.93E-01	×	×	-1.48E+02	0.987
	EM	2.62E-03	-1.84E-02	×	7.45E-01	×	-6.58E+01	0.992
	EF	3.19E-03	-1.07E-02	×	×	6.40E-01	-8.23E+01	0.978

Table 1 Coefficients of econometric models

Almost all the models' coefficients are statistically significant (the exception is the coefficient of the variable PD for EF model of 2011). Therefore, the influence of included explanatory variables on explained variables is significant. An additional advantage of the models is high value of determination coefficient R^2 – it provides a very good fit of the model to the empirical data. In the next step of the research, in order to examine the merits of taking into account the spatial factor in the description of the level of employment, the spatial autocorrelation of the residuals was examined. The obtained results are shown in table 2.

Year	ET model		EM model		EF model	
	I	p-value	I	p-value	I	p-value
2011	0.29	3.15E-06	0.22	2.38E-04	0.34	4.93E-08
2015	0.27	9.04E-06	0.22	2.51E-04	0.33	9.01E-08

Table 2 Moran's I statistics for models residuals

The calculated I Moran's statistics for the models residuals are positive and statistically significant. Statistically significant values of these characteristics mean that there is the phenomenon of spatial autocorrelation – therefore, we cannot speak of a random distribution of the model residuals. It can be observed that within five years the spatial dependence remained almost at the same level. In addition, obtained coefficients take the lowest values in the models for the number of employed men, while the highest listed are for models for the number of employed women (stronger spatial dependence is therefore observed for employment of women).

Subsequent analysis showed non-random residuals, both positive and negative – this situation tends to apply spatial methods in modelling of employment. The spatial lag model (SLM) and spatial error model (SEM) were considered. Diagnostics was carried out with a view to indentify the right model, and then the additional statistics of selected diagnostic tests were calculated. Table 3 shows the values of three information criteria, on the basis of which the right model is indicated. We have applied the Akaike criterion (AIC), the Bayesian criterion (BIC) and the LogLik criterion. The best model is the one for which the AIC and BIC values are the lowest, while for the LogLik criterion the values are the highest.

Model		Values of information criteria (2011)			Values of information criteria (2015)		
		AIC	BIC	logLik	AIC	BIC	logLik
ET	OLS	961.56	973.71	-475.78	931.38	943.54	-460.69
	SEM	**934.95**	**949.53**	**-461.47**	**910.61**	**925.20**	**-449.31**
	SLM	959.98	974.56	-473.99	933.10	947.69	-460.55
EM	OLS	820.74	832.90	-405.37	786.29	798.45	-388.15
	SEM	**801.53**	**816.11**	**-394.76**	**773.01**	**787.59**	**-380.50**
	SLM	820.01	820.01	-404.01	788.24	802.82	-388.12
EF	OLS	873.04	885.20	-431.52	847.68	859.83	-418.84
	SEM	**841.28**	**855.87**	**-414.64**	**818.52**	**833.11**	**-403.26**
	SLM	870.31	884.89	-429.15	848.14	862.72	-418.07

Table 3 The values of information criteria for models

The results indicate that in this issue spatial error model is the best one. The coefficients of the estimated models are shown in table 4. In this table, next to the parameters of the models, the autocorrelation coefficients values λ are presented which are at a high level.

Year	Explained variable	Explanatory variables					Intercept	λ
		GDP	PD	PT	PM	PF		
2011	ET	3.26E-03	-3.03E-02	6.74E-01	×	×	-7.99E+01	0.75
	EM	1.65E-03	-1.74E-02	×	7.33E-01	×	-4.19E+01	0.73
	EF	1.68E-03	-1.29E-02	×	×	6.15E-01	-4.02E+01	0.76
2015	ET	2.85E-03	-3.08E-02	7.04E-01	×	×	-8.11E+01	0.70
	EM	1.45E-03	-1.76E-02	×	7.56E-01	×	-4.08E+01	0.65
	EF	1.45E-03	-1.33E-02	×	×	6.52E-01	-4.16E+01	0.74

Table 4 Results of spatial error model estimation

The coefficients of each of the models and the autocorrelation coefficients are statistically significant. Table 5 shows also the values of statistics of selected diagnostic tests. The fact that the conditions are met, which should meet the properly specified model, confirms the correctness of the spatial model selection.

Model	Statistics values (2011)			Statistics values (2015)		
	Wald (λ)	LR_{ERROR}	LM_{ERROR}	Wald (λ)	LR_{ERROR}	LM_{ERROR}
ET	89.45	28.61	15.75	62.26	22.77	14.10
EM	78.03	21.22	9.01	44.00	15.28	8.95
EF	94.71	33.76	22.44	81.94	31.15	21.48

Table 5 Statistics values of selected diagnostic tests

3.3 Weights matrices and the quality of spatial models of employment

In the analyses six types of weights matrices were included. Three types were built on the basis of the contiguity criterion: row standardised first-order contiguity matrix (**W**), basic binary coding matrix (**B**), variance stabilizing matrix (**S**). The other included weights matrices are: *k*-nearest neighbours matrix ($k = 5$), matrix of neighbours within *d* km ($d = 200$), matrix based on the criterion of the inverse distance (inv_dist). It was examined, for which of the following weights matrix, spatial error models were the best. Table 6 shows the values of Akaike information criterion for the estimated models (the other two criteria yielded similar results).

Model	The type of weights matrix					
	W	**B**	**S**	$k = 5$	$d = 200$	inv_dist
ET (2011)	**934.95**	936.92	935.09	939.77	935.26	936.58
EM (2011)	**801.53**	803.99	802.19	804.91	804.17	803.94
EF (2011)	841.28	843.81	841.77	847.79	**841.13**	841.73
ET (2015)	**910.61**	912.26	910.94	913.41	911.68	911.84
EM (2015)	**773.01**	774.53	773.57	775.22	775.15	773.81
EF (2015)	**818.52**	821.28	819.26	822.57	818.95	819.55

Table 6 The values of Akaike information criterion for SEM models

Based on obtained values can be noted that in most cases, the best was the SEM model that includes row standardised first-order contiguity matrix (**W**), which is the most commonly used weights matrix. Therefore, models, whose coefficients are listed in table 4, in the best way describe the level of employment. An exception was the EF model of 2011 – in this case the lowest value observed for the matrix of neighbours within d = 200 km, however, the difference between the values of Akaike criterion of this model in comparison to the model with **W** matrix was small.

4 Conclusion

In the analyses linear spatial models were constructed based on data describing the level of employment (in total and by gender) in selected regions of Central Europe. At the beginning of the study, it was analysed whether using spatial modelling was justified. Each of the concerned models showed spatial autocorrelation. The analyses have shown that for all concerned dependences the best model is the spatial error one. Proper model specification was confirmed also by means of appropriate diagnostic tests.

Due to the fact that the type of spatial matrix may be of key importance in estimating the models, six types of matrices were included. The results of the analyses showed that, in the case of five of six considered models, the best is the one that includes the most common matrix – the row standardised first-order contiguity matrix. Only in one case a model with a different weights matrix was indicated (the difference between the values of the AIC of suggested model and the model with **W** matrix was very small).

The analyses showed that examining the level of employment (including selected variables and regions) it is appropriate to take into account the spatial modelling. Very good results in this case gives estimation of spatial error models including the row standardised first-order contiguity matrix.

References

[1] Anselin, L.: *Spatial analysis with GeoDa. 4. Spatial regression*. University of Illinois, Urbana Champaign, 2006.
[2] Anselin, L., and Bera, A.: Spatial dependence in linear regression models with an introduction to spatial econometrics. In: *Handbook of Applied Economic Statistics* (Ullah, A., and Giles, D. E. A., eds.). Springer-Verlag, Berlin, 1998.
[3] Cliff, A. D., and Ord, J. K.: *Spatial process: models and applications*, Pion, London, 1981
[4] Kopczewska, K.: *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R Cran*. CeDeWu, Warszawa, 2011.
[5] Mastalerz-Kodzis, A., and Pośpiech, E.: Application of spatial regression in employment characteristics modelling. *Acta Universitatis Lodziensis Folia Oeconomica* (2016) [in print].
[6] Mastalerz-Kodzis, A., and Pośpiech, E.: Spatial and temporal analysis of labour market characteristics. *Folia Oeconomica Stetinensia* (2016) [in print].
[7] Pietrzykowski, R.: Wykorzystanie metod statystycznej analizy przestrzennej w badaniach ekonomicznych. *Roczniki Ekonomiczne Kujawsko-Pomorskiej Szkoły Wyższej w Bydgoszczy* **4** (2011), 97–112.
[8] Pośpiech, E.: Analiza przestrzenna bezrobocia w Polsce. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach* **227** (2015), 59–74.
[9] Pośpiech, E.: Modelowanie przestrzenne charakterystyk rynku pracy. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach* **265** (2016), 69–79.
[10] Suchecki, B., eds.: *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*. C.H. Beck, Warszawa, 2010.
[11] Tobler, W.: A computer model simulating urban growth in Detroit region. *Economic Geography* **46(2)** (1970), 234–240.
Data: Eurostat http://ec.europa.eu/eurostat/web/regions/data/database (21.04.2017).

Line Integral in Optimal Control Problems

Pavel Pražák[1]

Abstract. Many problems encountered in management and economics can be formulated as optimal control problems. To solve an optimal control problem necessary conditions known as Pontryagin's maximum principle are introduced first. These conditions are formulated as a system of ordinary differential equations - either as an initial problem or as a boundary value problem - and they give us a basic idea about possible optimal solution to the given problem. The aim of this paper is to describe a class of optimal control problems that can be solved without using Pontryagin's maximum principle and without using a system of ordinary differential equations. At first a class of optimal control problems that can be formulated as a line integral is introduced. Then general results for finite time horizon problems that are based on Green's theorem are presented. Finally a particular use of the described method for neoclassical growth model with linear utility function on finite time horizon is introduced. The received solution corresponds with the solution acquired by Pontryagin's maximum principle.

Keywords: Green's theorem, growth model, line integral, optimal control.

JEL classification: C44, C61
AMS classification: 49J15, 49N90

1 Introduction

Dynamical systems with control variables play an important role in modern economics, [2], [8], [7] and in management as well, [9], [5], [6]. Especially models of growth theory are usually formulated as dynamic optimization problems, see [2]. In the class of one-dimensional optimal control problem there is an important subclass where singular paths can be found. For singular paths the particular control cannot be defined well, [5], [10], and there emerges a question how to characterize an optimal solution to the given problem. We will consider the approach proposed in [4] that consider that the objective function of the given optimal control problem can be changed into a line integral and then Green's theorem can be applied. We also present optimal control paths that keep turnpike property: on a fixed and a sufficiently long time horizon with the given initial and terminal states, the optimal path first starts from the initial state then reach a steady state, stay on this arc as long as possible and then leave it to reach the terminal state.

2 Optimal Control, Line Integral and Double Integral

In this section we consider the following class of one-dimensional optimal control problems over a given finite time horizon $T, T \in \mathbb{R}_+$,

$$\max(\mathcal{J}(x,u)|\ x \in M_1, u \in M_2), \tag{1}$$

where

$$\mathcal{J}(x,u) = \int_0^T \exp(-rt) g(x(t),u(t))dt \tag{2}$$

and

$$M_1 = \{x \in PC^1([0,T],\mathbb{R})|\ \dot{x} = f(x,u),\ x(0) = x_0,\ x(T) = x_T,\ u \in M_2\} \tag{3}$$

$$M_2 = \{u \in PC([0,T],\mathbb{R})|\ u(t) \in U = [u_1(x), u_2(x)]\}. \tag{4}$$

For optimal pair $(\hat{x}(t), \hat{u}(t))$ the condition $\mathcal{J}(\widehat{x},\widehat{u}) \geq \mathcal{J}(x,u),\ x \in M_1,\ u \in M_2$ is valid. It will be shown that the difference

$$\mathcal{J}(\widehat{x},\widehat{u}) - \mathcal{J}(x,u) \geq 0 \tag{5}$$

can be written as a line integral and moreover it can be written as a double integral. For this special case we need to consider that both real functions f and g have a special form. Following [9] it is particularly considered that

$$g(x,u) = A(x) + B(x)\Phi(x,u), \tag{6}$$

$$f(x,u) = a(x) + b(x)\Phi(x,u),\ b(x) \neq 0, \tag{7}$$

where all functions A, a, B, b and Φ are continuously differentiable with respect to x and u respectively.

[1]University of Hradec Králové, Faculty of Informatics and management, Department of Informatics and Quantitative Methods, Rokitanského 62, Hradec Králové, Czech Republic, pavel.prazak@uhk.cz

2.1 Transformation of the given problem

The given particular forms of functions f and g given by (6) and (7) allow us to transform the optimal control problem (1) – (4) into a more concise form. If (7) is substituted into the right side of differential equation at (3) we gain

$$\dot{x} = f(x,u) = a(x) + b(x)\Phi(x,u)$$

and therefore

$$\Phi(x,u) = \frac{\dot{x} - a(x)}{b(x)}.$$

Now we can rewrite (6) as

$$g(x,u) = A(x) + B(x)\Phi(x,u) = A(x) + B(x) \cdot \frac{\dot{x} - a(x)}{b(x)}$$

which means that

$$g(x,u) = M(x) + N(x)\dot{x}, \tag{8}$$

where

$$M(x) = A(x) - a(x)\frac{B(x)}{b(x)}, \quad N(x) = \frac{B(x)}{b(x)}. \tag{9}$$

The relation (8) allows us to rewrite objective function (2) when it is possible to write

$$\int_0^T \exp(-rt)g(x(t),u(t))dt = \int_0^T (\exp(-rt)M(x) + \exp(-rt))N(x)\dot{x})\,dt \tag{10}$$

If we put $v = \dot{x}$ as a new control variable and use right side of (10) as a new objective function, we can consider the optimal control problem (1) – (4) in the following form

$$\max(\mathcal{J}(x,v)|\ x \in M_1, v \in M_2), \tag{11}$$

where

$$\mathcal{J}(x,v) = \int_0^T \exp(-rt)\,(M(x) + N(x)v)\,dt \tag{12}$$

and

$$M_1 = \{x \in PC^1([0,T],\mathbb{R})|\ \dot{x} = v,\ x(0) = x_0,\ x(T) = x_T,\ v \in M_2\} \tag{13}$$

$$M_2 = \{v \in PC([0,T],\mathbb{R})|\ v(t) \in V = [v_1(x), v_2(x)]\}. \tag{14}$$

2.2 Line Integral

With the special form (6) and (7) of functions f and g the objective function $\mathcal{J}(\cdot)$ given by (2) and (10) can be rewritten as a line integral along a curve C in (t,x) space. Let

$$\mathbf{f}(t,x) = (f_1(t,c), f_2(t,x)) = (\exp(-rt)M(x), \exp(-rt)N(x)) \tag{15}$$

be the vector field and let

$$\varphi(t) = (t, x(t))$$

be the parametrization of a curve C in (t,x) space. Then

$$\dot{\varphi}(t) = (1,\ \dot{x}(t))$$

and $\mathbf{f} = \mathbf{f}(t)$. Instead of (10) it is possible to write

$$\int_0^T (\exp(-rt)M(x) + \exp(-rt))N(x)\dot{x})\ dt = \int_0^T \mathbf{f}(t) \cdot \dot{\varphi}(t)\,dt = \int_C \mathbf{f} \cdot d\mathbf{s}. \tag{16}$$

Let us finally consider that $(\hat{x}_{seg}, \hat{u}_{seg})$ and (x_{seg}, u_{seg}) are segments of optimal pair $(\hat{x}, \hat{u})$ and admissible pair (x,u) respectively that shape a closed, positively oriented and simple curve C, see Figure 1, then it is possible to write

$$\oint_C \mathbf{f} \cdot d\mathbf{s} = \mathcal{J}(\hat{x}_{seg}, \hat{u}_{seg}) - \mathcal{J}(x_{seg}, u_{seg}) > 0, \tag{17}$$

This is the reason why we need to deal with line integrals over closed curves in more details.

2.3 Double Integral

Green's theorem relates a line integral along a simple closed curve C in $\mathbb{R}^2$ to a double integral over the region D enclosed by the curve C, [3]. It is useful to remind that the closed curve C has the positive orientation if the region D is on the left side when we follow the curve C.

Theorem 1. *Let $C \subset \mathbb{R}^2$ be a positively oriented, piecewise smooth, simple, closed curve and let D be the region enclosed by the curve. Let $\mathbf{f} = (f_1, f_2)$ be C^1 vector field on an open set $O \subset \mathbb{R}^2$ containing D then*

$$\oint_C \mathbf{f} \cdot d\mathbf{s} = \int_D \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\, dx_1\, dx_2. \tag{18}$$

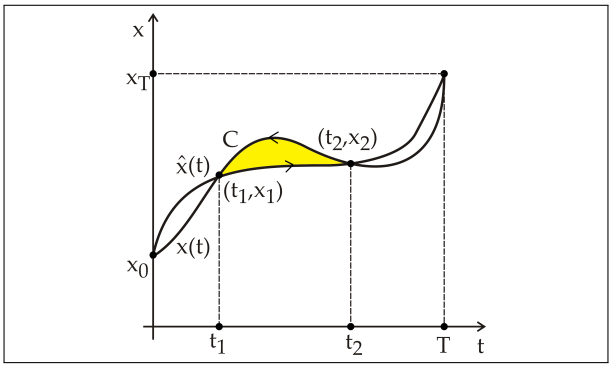


Figure 1 Positively oriented piecewise smooth, simple, closed curve C given by a segment of optimal trajectory $\hat{x}(t)$ and another admissible trajectory $x(t)$.

Now consider a vector field (15). For a positively oriented, piecewise smooth, simple and closed curve C in space (t, x), see Figure 1, the latter theorem can be applied to express the difference of objective functions (5) within some time interval as a double integral over the area D enclosed by the curve C. At first notice that

$$\frac{\partial f_2}{\partial t}(t,x) = -\exp(-rt) r N(x), \quad \frac{\partial f_1}{\partial x}(t,x) = \exp(-rt) M'(x).$$

It follows that

$$\oint_C \mathbf{f} \cdot d\mathbf{s} = \int_D \frac{\partial f_2}{\partial t}(t,x) - \frac{\partial f_1}{\partial x}(t,x)\, dt\, dx = \int_D -\exp(-rt)[M'(x) + rN(x)]\, dt\, dx \tag{19}$$

If we denote $I(x) = -[M'(x) + rN(x)]$, use (5), (2), (12) and (16) we can finally write

$$\oint_C \mathbf{f} \cdot d\mathbf{s} = \int_D \exp(-rt) I(x)\, dt\, dx. \tag{20}$$

2.4 Most Rapid Approach Path

Some optimal control problems have the following properties: It is optimal to stay on the stationary or singular path as long as possible or if the process starts out of the stationary or singular paths it is optimal to reach it in the fastest way. In other words stationary path $x(t) = x_s$ or singular path should be approach as fast as possible. These situation deserve to describe in more details. If $x_0 \neq x_s$ or $x_T \neq x_s$, where x_s is a stationary solution to the evolution equation of (1)–(4) then

$$x(t) = x_s,\ t \in [0,\, T]$$

is not a feasible path of the problem (1)–(4). For further discussion we need some new concepts, see [9].

Definition 1. Let $x \in M_1$ be a feasible path and let

$$\tau_1 = \min(t \in (0,T)|\ x(t) = x_s).$$

The *nearest approach segment from x_0 to x_s*, $\widehat{x}_1(t) = x(t)|_{t\in[0,\tau_1]}$, is defined as a segment for which the following condition is valid:

$$|\widehat{x}_1(t) - x_s| \leq |x(t) - x_s|,\ t \in [0, \tau_1]$$

for all feasible paths $x \in M_1$ Let $x \in M_1$ be a feasible path and let

$$\tau = \max(t \in (0,T)|\ x(t) = x_s).$$

The *nearest approach segment from x_s to x_T*, $\widehat{x}_2(t) = x(t)|_{t\in[\tau_2,T]}$, is defined as a segment for which the following condition is valid:

$$|\widehat{x}_2(t) - x_s| \le |x(t) - x_s|,\ t \in [\tau_2, T]$$

for all feasible paths $x \in M_1$

2.5 Optimal Solution

To describe an optimal solution to (1)–(4) with (6) and (7) we need to know where the integrand in (20) takes positive or negative values. To solve this problem we first consider the following algebraic equation

$$I(x) = -[M'(x) + rN(x)] = 0. \tag{21}$$

It can be shown that (21) characterizes a singular path of the problem (16)–(14). We denote by $\mathbf{K}$ the set of all solutions to (21)

$$\mathbf{K} = \{x \in \mathbb{R} | I(x) = 0\}.$$

In this paper we consider only the case when $\mathbf{K} = \{x_s\}$ is reduced to a set with exactly one element. Now we can show that the following theorem, cf. [9], for sufficiently long planning period is valid.

Theorem 2. *Let*
(i) x_s is the unique solution to (21) and x_s is feasible,
(ii) u_s is the unique solution to $f(x_s, u) = 0$, $u \in [u_1(x_s), u_2(x_s)]$
(iii)$\sigma(x)I(x) > 0$, where $\sigma(x) = sgn(x - x_s)$, $x \in \mathbb{R} \setminus \{x_s\}$,
(iv)there is a nearest approach segment $x_1(t), t \in [0, \tau_1]$ from x_0 to x_s,
(v) there is a nearest approach segment $x_2(t), t \in [\tau_2, T]$ from x_s to x_T

Then

$$\widehat{x}(t) = \begin{cases} x_1(t), & t \in [0, \tau_1], \\ x_s & t \in (\tau_1, \tau_2), \\ x_2(t) & t \in (\tau_2, T], \end{cases}$$

is optimal solution to (1)–(4) with (6) and (7).

It is necessary to emphasize that assumptions (iv) and (v) mean that the planning horizon $[0, T]$ of the optimal control problem is sufficiently long for nearest approach segments.

To prove the stated theorem we first consider an admissible trajectory $x(t)$ and a partition $(t_i)_{i=0}^n$ of $[0, T]$ such that $0 = t_0 < t_1 \cdots < t_n = T$ and for any i, $i \in \{0, \ldots, n-1\}$, either $x(t_{i+1}) = x_s$ or $t_{i+1} = T$. Now it is possible to consider regions

$$D_i = \{(t, x) \in \mathbb{R}^2 |\ t \in [t_i, t_{i+1}] \text{ and } x \in X_i(t)\},$$

where

$$X_i(t) = \begin{cases} [\hat{x}(t), x(t)], & \text{if } \sigma(x(t)) > 0, t \in [t_i, t_{i+1}], \\ [x(t), \hat{x}(t)], & \text{if } \sigma(x(t)) < 0, t \in [t_i, t_{i+1}]. \end{cases}$$

It means that the set $X_i(t)$ depends on whether $x(t) \le x_s$ or $x(t) \ge x_s$ for $t \in [t_i, t_{i+1}]$. Let us consider that the orientation of the boundary C_i of the region D_i is given by optimal control $\hat{x}$. Then C_i is a positively oriented closed curve if $X_i(t) = [\hat{x}(t), x(t)]$ and negatively oriented curve if $X_i(t) = [x(t), \hat{x}(t)]$. In other words C_i is positively oriented if $\sigma(x(t)) > 0$ and negatively oriented if $\sigma(x(t)) < 0$, see also Figure 1. If we use (20) then regardless of orientation of closed curves C_i it is possible to write

$$\oint_{C_i} \mathbf{f} \cdot d\mathbf{s} = \sigma(x) \int_{D_i} \exp(-rt) I(x)\, dt\, dx > 0,$$

for each i, $i \in \{0, 1, \ldots, n-1\}$. It means that

$$\mathcal{J}(\hat{x}, \hat{u}) - \mathcal{J}(x, u) = \sum_{i=0}^{n-1} \oint_{C_i} \mathbf{f} \cdot d\mathbf{s} > 0$$

which means that $J(\hat{x}, \hat{u})$ is optimal, cf. (5).

3 Neoclassical Growth Model with Linear Utility Function

To present how to apply the Green's theorem approach described in previous section 2, we shall consider a simple neoclassical growth model with linear utility function and finite time horizon T, $T \in \mathbb{R}_+$. The model describes an economic unit – like a state or a sector – which produces one type of goods. In compact form the model can be formulated as follows, for more details see [1],

$$\max(\mathcal{J}(k,c)|\ k \in M_1, c \in M_2), \tag{22}$$

where

$$\mathcal{J}(k,c) = \int_0^T \exp(-\beta t) U[c(t)]dt, \tag{23}$$

and U is a standard utility function, β, $\beta \in \langle 0,1)$, is a discount rate. Finally $c(\cdot)$ is per capita consumption. Further

$$M_1 = \{k \in PC^1([0,T],\mathbb{R})|\ \dot{k} = f(k) - (n+\delta)k - c,\ k(0) = k_0,\ k(T) = k_T,\ c \in M_2\} \tag{24}$$

where $k(\cdot)$ is per capita capital, that represents the state function of the problem and f is an intensive neoclassical production function. Moreover n, $n \in [0,1)$, is a growth rate of labour and δ, $\delta \in [0,1]$, is a rate of capital depreciation. The variable $c(\cdot)$ can be considered as the control input function of the problem and

$$M_2 = \{c \in PC([0,T],\mathbb{R})|\ c(\cdot) \in [0, f(k)]\}. \tag{25}$$

Since government can more likely influence investment than consumption, we reformulate the given problem for saving rate instead of for consumption. Particularly it is possible to consider that $c(\cdot) = (1 - s(\cdot))f(k(\cdot))$, where $s(\cdot)$, $s(\cdot) \in [0,1]$, is a saving rate. To be able to use the Green's theorem approach we shall consider that utility function is linear, i.e. $U(c) = c$. Thus the problem is to maximize the objective function

$$\mathcal{J}(k,s) = \int_0^T \exp(-\beta t)(1 - s(t))f(k(t))dt \tag{26}$$

subject to $\dot{k} = sf(k) - (n+\delta)k,\ k(0) = k_0,\ k(T) = k_T$ and $s \in [0,1]$. The given problem satisfies the special conditions (6) and (7). In particular we can observe that

$$A(k) = f(k),\ B(k) = -1,\ \Phi(s,k) = sf(k) \text{ and } a(k) = -(n+\delta)k,\ b(k) = -1.$$

If we substitute these functions into (9) and make some algebraic arrangement we gain

$$M(k) = f(k) - (n+\delta)k,\ N(k) = -1.$$

Now we can use (21) and find

$$I(k) = -[M'(k) + \beta N(k)] = -[f'(k) - (n+\delta+\beta)].$$

It is possible to show that there is exactly one k_s such that $I(k_s) = 0$, which means that $f'(k_s) = n + \delta + \beta$, for details see [1]. The existence of such k_s is the assumption (i) of Theorem 2. The stationary solution of the evolution equation is given by the equation

$$sf(k_s) - (n+\delta)k_s = 0. \tag{27}$$

Now we can find a stationary control

$$s_s = \frac{(n+\delta)k_s}{f(k_s)}, \tag{28}$$

which is the assumption (ii) of Theorem 2. Moreover it is possible to show, that $s_s \in (0,1)$. Let k_1 be the nearest approach segment from k_0 to k_s and similarly k_2 be the nearest approach segment from k_s to k_T. Now Theorem 2 can be used and the optimal path of capital $\hat{k}$ can be characterized as follows

$$\widehat{k}(t) = \begin{cases} k_1(t), & t \in [0,\tau_1], \\ k_s & t \in (\tau_1,\tau_2), \\ k_2(t) & t \in (\tau_2,T], \end{cases} \tag{29}$$

where τ_1 or τ_2 are appropriate instants within the interval $[0,T]$. The introduced optimal state path has the following property: we use such a control $\hat{s}$ policy to get on the stationary path k_s as quickly as possible stay there for most

of the planning time period and finally leave it as late as possible to reach the given terminal state k_T. For further analysis we assume that $k_0 < k_s$ and $k_T > k_s$. Then the following optimal control can be considered

$$\widehat{s}(t) = \begin{cases} 1, & t \in [0, \tau_1], \\ s_s & t \in (\tau_1, \tau_2), \\ 1 & t \in (\tau_2, T], \end{cases}$$

This part of discussion would deserve more details, but we put it off to another paper. If the initial and terminal state in neoclassical growth model are changed subject to the stationary state k_s we could find different solution. Nevertheless the character of all solutions are similar - in the middle of the optimal trajectory there is a stationary part.

4 Conclusion

The analysis described in this paper is mainly based on the assumption of sufficiently long and finite time horizon of optimal control problems. We have showed that in this instance we can use a very simple control that uses at least three values - minimal or maximal possible values or a singular value. In the future work we would like to deal with both the short planning period and the very long planning horizon which is usually modeled by infinite time horizon, see also [7]. It would be also very useful to find or at least to specify instants τ_1, τ_2 that characterize nearest approach segments. We consider that it coul be done by solving another optimal control problems.

Acknowledgements

Support of the Specific research project of the Faculty of Informatics and Management of University of Hradec Králové and the technical help of student Kateřina Frončková is kindly acknowledged.

References

[1] Brunovský, P.: *Mathematical theory of optimal control*, (in Slovak), Alfa, Bratislava, 1980.

[2] Chiang, A.C.: *Elements of dynamical optimization,* Wavelend Press, Illinois, 2000.

[3] Marsden, J.E., Tromba, A.J.: *Vector Calculus*. W. H. Freeman and Comp., New York, 1996.

[4] Miele, A.: Extremization of linear Integrals by Green's Theorem, *Optimization Technique,* ed. by G. Leitmann, Academic Press, New York, 1962.

[5] Kamien, M.I., Schwartz, N.I.: *Dynamic Optimization, the Calculus of Variations and Optimal Control in Economics and Management*. North-Holland, Amsterdam, 1991.

[6] Pražák, P.: On the Vidale - Wolfe Advertising Model, In: *Mathematical Methods in Economics 2004* (ed. Bauer, L.), MU Brno, 2004, 269 - 274.

[7] Pražák, P.: On Necessary Transversality Condition for Infinite Horizon Optimal Control Problems, In: *Mathematical Methods in Economics 2011- part II* (ed. Dlouhy, M.), Professional Publishing, Prague, 2011, 575-580.

[8] Seierstad, A., Sydsæter, K.: *Optimal Control Theory with Economic Application,* Elsevier, Amsterdam, 1987.

[9] Sethi, S.P.: Nearest feasible paths in optimal control problems: theory, examples, and counterexamples. *Journal of Optimization Theory and Applications* **23** (4) (1977), 563–579.

[10] Sethi, S.P., Thompson, G.L.: *Optimal Control Theory, Applications to Management Science and Economics*. Kluwer Academic Publishers, Boston, Dordrecht, London, 2003.

Tradeoff between Economic and Social Sustainability of Bus Network

Vladimír Přibyl[1], Jan Černý[2], Anna Černá[3]

Abstract. The social pillar of bus network sustainability is mainly affected by the accessibility of services. The same role is played by costs for the economic pillar. The tradeoff between these two factors consists in looking for the bus network with minimal costs maintaining the acceptable level of accessibility.
The paper presents a mathematical model of the network of roads with sidewalks for pedestrians. Each vertex is a "candidate" for a bus stop and it is given the number of demanding passengers at it and for each edge it is given the generalized cost, if it is chosen to the bus subnetwork. Moreover, it is given the maximum acceptable accessibility limit (i.e. the limit of the average walking distance from the vertices not chosen to the subnetwork to the closest stops). The problem is to find such a bus subnetwork that does not exceed the accessibility limit and minimizes the total cost.
Three methods are proposed and verified: Two exact, i.e. the Depth-first-Search and Linear Programming and one heuristics.

Keywords: Bus, transport, subnetwork, sustainability, social, economic.

JEL Classification: C44, C61, R42
AMS Classification: 90B10, 90C11

1 Introduction

In general, sustainability has three pillars: environmental, economic and social (see e.g. Litman [3]). For the bus transport, the tradeoff between the last two is a crucial issue. While the strength of social pillar is mainly affected by the accessibility of services, the same role is played by costs for the economic pillar. The tradeoff between these two factors consists in looking for the bus network with minimal costs maintaining the acceptable level of accessibility. The same can be said for other modes of public transport, e.g. trolleybus or tram transports.

The notion accessibility is meant basically in the geographic sense – the lower average walking distance, the higher accessibility. However, there are two other features of accessibility for traveler, together with two features of the traffic supply, which affect its accessibility for travelers. The first is the time – a traveler needs to get from the starting point to the destination in a reasonable time. The second is the fare – the passenger does not intend, and often is not even able, to pay an exorbitant price for travelling.

The passenger loses time in three ways – walking, waiting and riding. Walking time is proportional to the walking distance and it is already taken into account. Waiting time and riding time depend on the route. The straighter route, the shorter riding time and moreover, the shorter headway, since the vehicles turn more quickly. And, by the way, the straighter route, the smaller fare. The resume is: the straighter routes, the higher accessibility.

The paper studies the following situation: A new (or strongly innovated) public transport system is to be introduced into the given network G of "roads". For the purposes of this text, the term "road" includes both off-highway and street roads. Regardless of whether public transport has operated on the existing network G, the introduction of the new system requires expensive building adjustments for all sections to be operated. Due to the budget limitations, it will not be possible to rebuild the entire existing network for the new system, but only its carefully selected part. The procedure for this selection is described in the later parts of the paper.

Generally, it is not new to look for the "cheapest" subgraph, but its required properties are different from the ones dealt in the current paper. Safari [4] required a subgraph of given connectivity and Vassilevska [5] seeked a subgraph isomorphic to another given graph

[1] University of Economics Prague, Faculty of Management, Jarošovská Street, Jindřichův Hradec, Czech Republic, e-mail: pribyl@fm.vse.cz.
[2] University of Economics Prague, Faculty of Management, Jarošovská Street, Jindřichův Hradec, Czech Republic, e-mail: cerny@fm.vse.cz.
[3] University of Economics Prague, Faculty of Management, Jarošovská Street, Jindřichův Hradec, Czech Republic, e-mail: cerna@fm.vse.cz.

2 Models of the Current and the New Networks and their Problems

2.1 Description of the Current Network

The current "old" network can be precisely described as a *finite strongly connected* (non-oriented) *graph* $G = (V, E, q, d, c)$ of roads (with sidewalks for pedestrians on the sides) where

the *vertex set* $V = \{1, \ldots, n\}$ represents the *passenger origins* or *destinations* and, simultaneously, *candidates for location of stops*; usually it is assumed that these points are close to intersections and therefore there is no need to differentiate them for managerial decision making;

the edge *set* E represents *direct links between neighboring vertices*;

the *function* q: $V^2 = V \times V \to \langle 0; \infty)$ represents *passenger demand*: $q(i, j)$ expresses the total average number of passengers from the vertex $i \in V$ to the vertex $j \in V$ (e.g. during one day), i. e. $Q = (q(i, j))$ is an O-D matrix of passenger demand; the denotations $q^+(i)$ and $q^-(j)$ mean the i^{th} row and j^{th} column sum of the matrix Q respectively, and $q^T = q^+(1) + \ldots + q^+(n) = q^-(1) + \ldots + q^-(n)$ is the total demand, i. e. the number of all demanding passengers;

the *function* d: $V^2 \to \langle 0; \infty)$ represents *distances* on G: $d(i, j)$ expresses the distance from the vertex $i \in V$ to the vertex $j \in V$, and, particularly, $d(e)$ is the length of the edge $e \in E$; the denotation $d(k, V') = \min\{d(k, j): j \in V'\}$ for all $k \in V$ and $V' \subset V$ means the distance of a vertex k from the vertex subset V',

the *function* c: $E \to \langle 0; \infty)$ represents *costs*: $c(e)$ expresses the costs induced by the choice of the edge $e \in E$ for the public transport (e.g. reconstruction of road surface, construction of trolleybus wires, tram rails etc.).

2.2 Requirements for the New System

The new system will operate on a subnetwork $G' = (V', E', q', d', c')$ of G, i. e. $V' \subset V$, $E' \subset E \cap V'^2$, $q'(i, j) = q(i, j)$ for all $i \in V'^2$, $c'(e) = c(e)$ and $d'(e) = d(e)$ for all $e \in E'$, and d' is the corresponding distance function on V'^2. It will use transport routes, connecting the *most important nodes* in G (health, business, education, sports and cultural centers, offices, railway stations, airports, ports etc.) chosen by local authorities, usually from the ones with relatively high values of demand. The set of them is denoted W and it is assumed $W \subset V$. The nodes from W are considered *candidates for terminals* in the new system. A set $W' \subset W^2$ is also chosen by local authorities from the pairs $(v, w) \in W^2$ s. t. $q(v, w)$ is relatively high among the elements of the O-D matrix Q.

The route r connecting the selected pair $(v, w) \in W'$ is said η*-feasible* for the given value $\eta \in \langle 1; \infty)$ if the length $d(r)$ of the route r meets the constraint $d(r) \leq \eta.d(v, w)$. The set of all η-feasible routes connecting the pair (v, w) is denoted $R(v, w\,|\eta)$. The set R of routes connecting pairs from W' is said η*-feasible* if $R \cap R(v, w\,|\eta)$ is a one-element set for each $(v, w) \in W'$. The most common values of η in practice are 1.0, 1.2 and 1.5. In the new system in future, an η-feasible set R of routes connecting pairs from W' may represent candidates for new bus lines, operating in the area of G.

The (geographic) *inaccessibility* of a subset V' on G equals to the average walking distance from the closest stop and it is expressed as follows:

$$g(V') = \frac{1}{q^T} \sum_{k \in V} \left(q^+(k) + q^-(k)\right) d(k, V') \tag{1}$$

Let R be an η-feasible set of routes connecting pairs from W', let $V(r) = \{i \in V: i \text{ is traversed by } r\}$, let $E(r) = \{e \in E: e \text{ is traversed by } r\}$ and let

$$V(R) = \bigcup_{r \in R} V(r)\,,\ E(R) = \bigcup_{r \in R} E(r)\,,\ c(E(R)) = \sum_{e \in E(R)} c(e) \tag{2}$$

The basic requirement for the new system is that for the given parameters $\lambda \in (0; \infty)$ and $\eta \in \langle 1; \infty)$ there exists an η-feasible set R of routes connecting pairs from W' s. t. $g(V(R)) \leq \lambda$ and $c(E(R))$ is minimum possible. The new system then operates on the subnetwork $G' = (V', E', q, d', c)$ where $V' = V(R)$, $E' = E(R)$.

2.3 Optimization Problem OP

The *optimization problem* **OP** is the following:

Given the network $G = (V, E, q, d, c)$, the set of important vertices $W \subset V$, the set of important pairs W', the O-D matrix Q and the parameters λ and η. The problem is to find such η-feasible set of routes R connecting pairs from W' that $g(V(R)) \leq \lambda$ and $c(E(R))$ is minimum possible.

If an η-feasible set of routes R connecting pairs from W' meets the constraint $g(V(R)) \leq \lambda$, then it is called the (η, λ)-*feasible solution* of the problem **OP**. Moreover, if an (η, λ)-*feasible solution* of the problem **OP** minimizes the value $c(E(R))$, then it is called *optimal*.

Any solution of **OP** can be characterized by three parameters: λ (expressing accessibility – social pillar of sustainability), $c(E(R))$ (costs – economic pillar) and η (both pillars). If for a practician, i. e. usually a manager of local public administration, the solution seems unbalanced, then, first, one can try to find a tradeoff changing η which will probably cause the opposite change of λ.

If no such solution is satisfactory for practicians, the reason can be expected in the fact that not negligible sources/sinks of passenger movements are too far from the candidate routes from the set R. Then one can try to extend the set W' and if it does not work then to rethink the possible extension of then whole set W.

3 Solution of the Problem OP

In this chapter, two exact methods of solution for the problem **OP** are presented: linear programming and depth-first-search technique (briefly DFS; for more details, see e.g. Kleinberg [6]).

3.1 LP solution of OP

Here, two different LP models for the solution of OP are described:

- LPP1 which works fast but often leads to (η, λ)-unfeasible solution of the problem **OP**;
- LPP2 which is much slower but very promising in the search for optimal solution of **OP**.

Both models work with the following binary variables:

x_e for each $e \in E$; $x_e = 1 \Leftrightarrow e$ is chosen to E',

y_r for each $r \in R(v, w)$, $(v, w) \in W'$; $y_r = 1 \Leftrightarrow r$ is chosen to R i. e. for future service between v and w,

z_i for each $i \in V$; $z_i = 1 \Leftrightarrow i$ is chosen to $V' = V(R)$,

s_{ij} for each $i \in V$, $j \in V'$; $s_{ij} = 1 \Leftrightarrow$ for the vertex i the vertex j is the closest in V'.

The first problem **LPP1** is to find the variables x_e and y_r s. t. the following constraints hold:

$$\sum_{r \in R(v,w)} y_r = 1 \text{ for each } (v, w) \in W' \tag{3}$$

$$x_e \leq \sum_{(v,w) \in W'} \sum_{r \in R(v,w)} \sum_{e \in r} y_r \leq M x_e \text{ for each } e \in E \text{ where } M \text{ is a great number (e. g. } 10^5) \tag{4}$$

$$C(E') = \sum_{e \in E} c_e x_e \rightarrow \min \tag{5}$$

Obviously, the solution $R = \{r: y_r = 1\}$ of the problem **LPP1** is η-feasible. Consequently, if it also meets the constraint $g(V(R)) \leq \lambda$ then it is (η, λ)-feasible and because of (5) it is an optimal solution of **OP**.

Otherwise, if the solution of **LPP1** does not meet $g(V(R)) \leq \lambda$ then the inaccessibility is too high and one has to look for some tradeoff between the values $g(V')$, $C(E')$ and η, where the first belongs to the social pillar of sustainability, the second to the economic pillar and the third affects both pillars. Solution of the following problem **LPP2** can help in it.

Find the variables x_e, y_r, z_i and s_{ij} s. t. the following constraints hold:

$$\sum_{r \in R(v,w)} y_r = 1 \text{ for each } (v, w) \in W' \tag{3}$$

$$x_e \leq \sum_{(v,w) \in W'} \sum_{r \in R(v,w)} \sum_{e \in r} y_r \leq M x_e \text{ for each } e \in E \tag{4}$$

$$x_{(i,j)} \leq \frac{z_i + z_j}{2} \text{ for each } (i, j) \in E \tag{6}$$

$$\sum_{j \in V} s_{ij} = 1 \text{ for each } i \in V \tag{7}$$

$$\sum_{j \in V} d(i, j) s_{ij} \leq d(i, k) z_k + M(1 - z_k) \text{ for each } i \in V, k \in V \tag{8}$$

$$g(V') = \frac{1}{q^T} \sum_{i \in V} \sum_{j \in V} d(i, j) q(i) s_{ij} \leq \lambda \tag{9}$$

$$\sum_{e \in E} c_e x_e \rightarrow \min \tag{5}$$

It may often happen that despite the result $g(V') > \lambda$ of **LPP1** the problem **LPP2** is resolvable. Than its solution resolves **OP** as well.
However, the opposite is possible too and thus a tradeoff is necessary. It may be done by the choice of new values of parameters η and λ, etc. as described at the end of 2.3.

3.2 DFS Technique for OP

From structural point of view, the use of DFS (depth first search) is simple. Before it starts, all sets $R(v, w \mid \eta)$ are constructed. Then DFS passes through the solution tree with the sets R as leaves. Each R having the property that $R \cap R(v, w \mid \eta)$ is a one-element set for each $(v, w) \in W'$.

4 Modification of OP

When **OP** is unsolvable or when its first solution is not acceptable for practicians, then the tradeoff seeking procedure, described at the end of 2.3, takes place. However, sometimes it does not reach an acceptable solution as well or the procedure itself may seem to be very subjective, depending of the choices "at the discretion". The following modification of **OP** tries to overcome these objections.

4.1 Problem $\mathbf{OP_{mod}}$

The *modified optimization problem* $\mathbf{OP_{mod}}$ is the following:

Given the network $G = (V, E, q, d, c)$, the set of important vertices $W \subset V$, the set of important pairs W', the O-D matrix Q and the parameters λ and η. The problem is to find such subnetwork $G' = (V', E', q, d', c)$ that

$$W \subset V' \tag{10}$$

$$g(V') = \frac{1}{q^T} \sum_{k \in V} \left(q^+(k) + q^-(k)\right) d(k, V') \leq \lambda \tag{11}$$

$$d'(e) = d(e) \text{ for all } e \in E' \tag{12}$$

$$d'(v, w) \leq \eta . d(v, w) \text{ for all } (v, w) \in W' \tag{13}$$

$$c(E') = \sum_{e \in E'} c(e) \rightarrow \min \tag{5}$$

The subnetwork $G' = (V', E', q, d', c)$ fulfilling the constraints (10)-(13), (5) is *optimal solution* of $\mathbf{OP_{mod}}$, if G' fulfils (10)-(13), then it is said *feasible solution* of $\mathbf{OP_{mod}}$.

Comparing with **OP**, the problem $\mathbf{OP_{mod}}$ does not require that $e \in E'$ only if e is traversed by a route from R, and therefore in the new system bus lines without terminals in W are allowed.

The authors have not found any suitable LP model for $\mathbf{OP_{mod}}$ yet. On the other hand, they succeed in use of DFS and, moreover, they found a well working heuristics.

4.2 DFS for $\mathbf{OP_{mod}}$

This DFS passes through the solution tree with the sets $E' \subset E$ as leaves with E as the root and restricted to such sets E' that each $w \in W$ is incident with some $e \in E'$. The neighboring leaf down is reached by omitting some edge $e \in E'$.

4.3 Heuristics for OP_{mod}

Since the following method is a heuristics, it guarantees only the feasibility, not optimality of solution. However, the difference between the achieved value of $c(A')$ and the actual minimum is usually small and sometimes zero.

0° Step: Put $G'' = G$, $B = \varnothing$,

1° Step: Find the set P'' of shortest paths $p(v, w)$ from v to w on G'' for each pair $(v, w) \in W^2$, $v \neq w$. If all paths from P'' contain only the arcs from the set B, then put $V'' = \{k \in V: k \in p(v, w)$ for some $p(v, w) \in P''\}$ and goto 2° Step.

If not, then put $B = B \cup \{a''\}$ where $a'' \in A - B$ s. t. a'' is traversed by maximum number of paths from P''. Put $G'' = (V, A, q, d'', c)$, where $d''(a) = d(a)$ for $a \in A - B$ and $d''(a) = d(a)/\eta$ for $a \in B$. Goto 1°Step.

2° Step: If (3) holds, i.e. $g(V'') \leq \lambda$, then $G' = (V'', B, q, d', c)$ is a feasible solution of **OP** with the minimum of $c(A')$ achievable by the heuristics (but not necessarily the absolute minimum reachable by an exact method). G' is the final result.

If (3) does not hold then find $w \in V - V''$ s. t.

$$g(V'' \cup \{w\}) = \min\{g(V'' \cup \{v\}): v \in V - V''\} \tag{14}$$

and find $w_1 \in V''$, $w_2 \in V''$ s. t. $(w_1, w) \in A$, $(w, w_2) \in A$ and

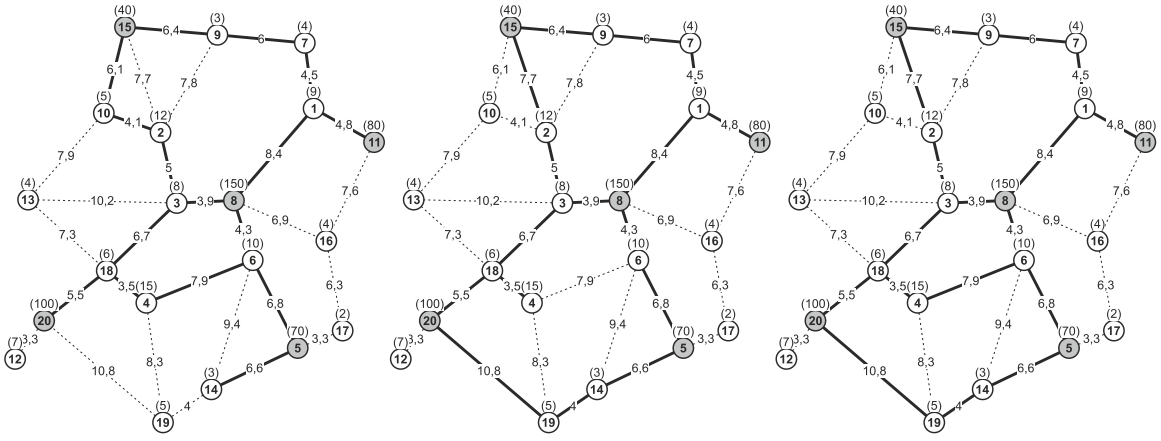
$$\begin{aligned} &d(w_1, w) + d(w, w_2) = \\ &\quad = \min\{d(v_1, w) + d(w, v_2): v_1 \in V'', v_2 \in V'', (v_1, w) \in A \text{ and } (w, v_2) \in A\} \end{aligned} \tag{15}$$

Then put $V'' = V'' \cup \{w\}$, $B = B \cup \{(w_1, w), (w, w_2)\}$ and goto 2° Step

5 Computational Experience

All above mentioned methods was implemented and tested on the randomly generated test network with 20 vertices and 30 edges. This network scope allowed to use exact methods with acceptable computing times. The topology of this network can be seen on any part of Figure 1. The numbers in the middle of edges are their lengths. Assuming a proportional relation between the lengths and costs of edges we put the costs equal to the lengths for testing purpose. The number in parentheses are the total demands of passengers in appropriate vertices. The important vertices (with high demand) are grayed out. Total costs of the original network were 191.3.

Tests were carried out for all combinations of different parameters λ and η and for different number of important pairs from the set {20-11; 15-5; 15-11; 15-20; 20-5}. Unfortunately, there is not enough space for presenting all results but in summary we can say, that the results of DFS and LP method for the problem OP were the same in all cases. The results of the heuristic method for OP_{mod} were not the optimal ones reached by DFS method in most cases, but the maximal relation between the optimal solution and the heuristic one was 1.17 (average 1.06).



DFS (costs 90.5, OP_{mod}) Heuristics (costs 94.9, OP_{mod}) LP and DFS (costs 102.8, OP)

Figure 1 The solutions of OP and OP_{mod} problems for $\lambda = 0.2$, $\eta = 1.2$ and complete set of 5 important pairs reached by different methods

Figure 1 depicts the resulting subnetworks in one particular case ($\lambda = 0.2$, $\eta = 1.2$ and complete set of 5 important pairs). It can be seen, that the solution reached by heuristics is slightly worse than the optimal one reached by DFS method. The OP is different than the OP_{mod} so it is not possible directly compare solutions of this problems, but one can see, that in case of the solution of OP there is no vertex in V' with the degree 1 except the vertices from the important pairs. This is the reason why it is not possible to reach total costs as low as in case of OP_{mod} solution in most cases. The lower is η and/or λ the higher is the difference.

Table 1 brings the overview of resulting costs of subnetworks reached as a solution of OP in case of complete set of 5 important pairs. It can be seen, that with increasing η and/or increasing λ is it possible to reach better solution (from the total costs point of view). Respectively, it is possible to look for the tradeoff between the economic and social pillars.

η	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 1$
1.0	U	U	U	91.4
1.1	U	U	115.2	91.4
1.2	U	122.7	102.8	81.4
1.3	U	108.4	87	70
1.5	118	103.5	81.3	53.1

Remark: U – there does not exist any feasible solution for this combination of λ and η

Table 1 Total costs of resulting networks (solutions of OP with complete set of 5 important pairs)

The computational times for the heuristics were less than 1 second in all cases. DFS methods takes significantly more time (in some cases several hours) and they are not suitable for larger networks. LP methods were quick even with the freeware LP solver LP Solve. In case of LP the main portion of time takes generating the sets $R(v, w|\eta)$, but the set $R(v, w|\eta)$for certain η is suitable not only for different λ, but for all smaller η.

6 Conclusion

The paper presented the problem how to choose a subnetwork for a new system of public transport, connecting the important nodes in a current system of roads that are not suitable for the new system and the budget limitations do not allow to reconstruct them all. This choice has to be a tradeoff between social aspect represented by accessibility requirement and the economic one represented by costs. The problem was formulated and solved mathematically by means of computer, using two types of exact methods and one heuristics.

The future research may be expected focused on the development of LP models embodying wider possibilities of practical requirements.

Acknowledgements

Supported by FM VŠE Project No. IP 632182.

References

[1] Černá, A., and Černý J.: *Manažerské rozhodování o dopravních systémech* (= *Managerial Decision Making on Transportation Systems.* A monograph in Czech) . Univerzita Pardubice, Pardubice, 2014.
[2] Černá, A. Černý J., Přibyl, V.: Bus Route Design in Small Demand Areas. Transport 26, 3(2011), 248-254.
[3] Litman, T., Burwell, D,: Issues in sustainable transportation. Int. J. Global Environmental Issues 6, 4(2006), 331-347.
[4] Safari, M.A. and Salavatipour, M.R.: A Constant Factor Approximation for Minimum λ-Edge-Connected k-Subgraph with Metric Costs. In: A. Goel et al. (Eds.): *APPROX and RANDOM 2008, LNCS 5171*, Springer-Verlag Berlin Heidelberg: 233–246.
[5] Vassilevska, V., Williams and R.,Yuster, R.: Finding the Smallest H-Subgraph in Real Weighted Graphs and Related Problems. In: M. Bugliesi et al. (Eds.): *ICALP 2006, Part I, LNCS 4051*, Springer-Verlag Berlin Heidelberg: 262–273.
[6] Kleinberg, J.; Tardos, É.: *Algorithm Design.* Pearson/Addison Wesley, Boston, 2006 , pp. 92–94

The concept of mixed method study for risk assessment in manufacturing processes

Angelina Rajda-Tasior[1]

Abstract. According to the latest edition of ISO 9001:2015 organizations must estimate and monitor the risk in each process and make a plan to minimize its impact. A practical problem is to choose the appropriate statistical method which will be adequate for the industry and subsequently allows for a reduce the risks involved. The problem is complex because there is lack of risk assessment methods which allow to prepare a full risk analysis, in particular dedicated for the industrial areas.
The article provides a concept of application of mixed method study for assessing the risk for one foundry in Poland. Castings production is an specific industry where the processes are very complicated and unstable, therefore the use of a multiple tool to illustrate the risk will be an added value for the company. This kind of approach allows to make a full risk assessment analysis what cannot be done by a single statistical techniques with some limitations.In this paper empirical and simulation data is used. The proposal of multi-tool can enable effective risk estimation, assist in determining the significance of the data set and to facilitate accurate decision-making. It can be used to predict and prevent problems, reduce cost and shorten the development time. The economic and practical benefits resulting from the application of the proposed solution are presented. The concept of mixed methods allows for the construction of more sensitive tool as well as a better broader understanding of the phenomena of risk in processes.

Keywords: ISO standards, risk analysis, quality assurance, semi-qualitative methods, multiple criteria decision making, simulations.

JEL Classification: C44, M11
AMS Classification: 62C07

1 Introduction

One of the key change in the 2015 revision of ISO 9001 is to establish a systematic approach to risk, rather than treat it as a single component of a quality management system. By taking a risk-based approach, an organization becomes proactive rather than purely reactive, it prevents or reduces undesired effects and promotes continual improvement. ISO 9001:2015 defines risk as the effect of uncertainty on an expected result and it can be defined by two parameters*: severity* and *probability*. Risk is defined as the combination of the probability of occurrence of risk and the severity of that risk [4] but also it is an influence on the goals of company [9].

2 Research problem description

The methodology of the risk management is effectively applied in many areas of trade and government [2]. There are many techniques of the risk assessment for finance, occupational safety, public health, insurance, emergency but there are only few dedicated for industries. In production there are some examples of the use of quality risk management today, but they are limited and do not represent the full contributions that risk management has to offer. Risk analysis and management depend mainly on intuition, judgment and experience. Moreover, each organization might recognize different potential problems, a different probability on each problem occurring and attribute different severities to them. Formal risk analysis and management techniques are rarely used due to a lack of knowledge and to doubts on the suitability of these techniques.

There are some documents for risk assessment suggested in the literature. The standard, ISO 31000:2009 is recommended as a basis for the most universal and comprehensive use in organizations which are facing a broadly conceived risk on a daily basis. The standard is a group of principles, framework and processes for use by any organization. It does not provide ready solutions but only the components of risk management which can be adopted. The standard ISO 31010:2009 is a set of methods and techniques recommended for the risk assessment in many areas in a general sense [10]. The basic vocabulary to develop common understanding of the risk management concepts and terms among organizations and functions, and across different applications and types can be found in a Guide [11]. Two methods for quality risk management can be found in Measurement

[1] University of Economics, Department of Statistics, Katowice, Poland, angelina.rajda@gmail.com

System Analysis Reference Manual [3]. There, the risks of making wrong or inconsistent decisions can be evaluated.

The problem when the risk is assessed it is necessary to take into account the appropriate method to evaluate its results. Moreover the output of a risk assessment can be a quantitative estimation of risk or a qualitative description of a range of risk. Risk can be expressed using qualitative descriptors, such as *high*, *medium*, or *low*. Sometimes the *risk score* methodology is used to further define descriptors in risk ranking. In the quantitative risk assessments, the risk estimation provides the probability of a specific consequence, given a set of risk which are generating circumstances. The quantitative risk estimation is useful for one particular consequence at a time. Alternatively, some risk management tools use a relative risk measure to combine multiple levels of severity and probability into an overall estimation of relative risk. The comparison is given by Table 1.

Techniques	Advantages	Disadvantages
Qualitative	- relatively simple and easy - relatively fast way of obtaining results - relatively subjective and optional - based on good practice and experience - allows for diversity of approaches - allows for identification of the risks and gives fully lists of risks (or ranking of risks) - can gives lots of information without financial impact and likelihood - good flexibility and possibility for modification	- gives limited differentiation between the risk levels - it is inaccurate - does not allows numerical aggregation with risk interactions and correlations - no possibility to generalize the results to the population - it is a subjective analysis and interpretation of results - it is the risk of misinterpretations and wrong conclusions - at times gives a low degree of relevancy
Quantita-tive	- allows to compare investigated phenomena and results - allows to determine the significance of differences between variables - allows numerical aggregation taking into account risk interactions and correlations - allows to express the results of estimating as a cost - permits cost-benefit analysis of risk response options - obtains knowledge based on objective data - estimated results are objective	- possibility of a faulty application methods - it can be time-consuming and costly - all measures must be defined - use of numbers may imply calculating errors - assumptions may not be apparent

Table 1. Risk assessment techniques comparison

3 The risk assessment process

Let the individual risk R_i define as follows:

$$R_i = L_i\, p(L_i) \tag{1}$$

where L_i will be the product of potential losses and $p(L_i)$ will be their probabilities. Than total risk will be the sum over individual risks (formula 2):

$$R_{total} = \sum_i L_i\, p(L_i) \tag{2}$$

In production area popular approach for risk assessment is risk indicator r_i according to following formula:

$$r_i = O_i L_i\, p(L_i) \tag{3}$$

where O_i is the possibility of detection of the problem.

Risk assessmentconsists of the three processes: identification of hazards, the analysis and evaluation of risks associated with exposure to those hazards (Figure 1). Quality risk assessments begin with a well-defined problem description. When the risk in question is well defined and an appropriate risk management tool is taken, the risk will be more readily identifiable.

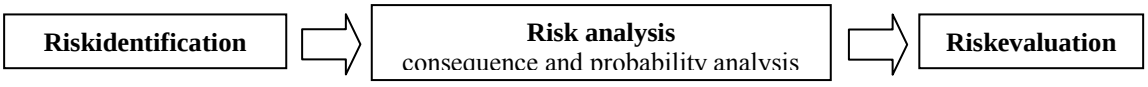


Figure1 Risk assessment procedure

The main purpose of the risk assessment is to establish possibly *wide list of risks* that will affect the defined objectives, regardless of whether their source is under the control of the organization, or it does not affect their existence.

Definition 1.Risk identification *is a systematic use of information to identify hazards referring to the risk description.* Information can include historical data, theoretical analysis, informed opinions, experience, data from complaints analysis et cetera. The process of risk identification includes identifying the possible consequences. This provides the basis for further steps in the quality risk management process, taking into account the cascade effect, which affects the appearance of the next risks.

Definition 2. Risk analysis *is the estimation of the risk associated with the identified hazards. It is the qualitative or quantitative process of linking the probability of occurrence and severity of risks.* In some risk management techniques, it is the ability to detect (O_i) - the possibility of detection of the problem (formula 3).

Definition 3. Risk evaluation *compares the identified and analyzed risk with against given risk criteria in order to determine whether or not a specified level of risk is acceptable or tolerable.* This process can be compared to idea of Shewhart control charts. The use of the control chart mainly illustrates the level of the risk. If the risk of the process is greater than assumed that the control limits are exceeded. Therefore, the task of the control charts in the aspect of risk management is to identify and to signal the risky processes in order to minimize risk.

4 Risk assessment methods. The analysis and verification

The awareness of the practical implications of risk quality and the need for its measurement can be monitored using 7 tools of statistical quality management [13, 16]:cause and effect diagram, check sheet, control chart (for example: acceptance control charts, control charts with arithmetic average and warning limits, cumulative sum charts, Shewhart control charts or weighted moving average) [5, 6, 7, 8]. The multivariate charts [14], histogram, Pareto chart, scatter diagram and flow chart also can be used.All of these tools can be useful in the risk assessment. They allow to control the quality on-line by the sequential elimination of the causes of excessive variability and thus reduce the risk of quality.

Techniques	**Risk assessment process**								
	Risk analy-sis								
	Risk detection	Consequences	Probability	Risk level	Risk evaluation	Difficulties level (1-5)	For most industries	Quantitative type	Qualitative type
Checklist	+	-	-	-	-	1	+	-	+
Preliminary Hazard Analysis	+	-	-	-	-	2	+	+	+
Brainstorming	+	-	-	-	-	1	+	-	+
Delphi Technique	+	-	-	-	-	1	+	+	+
SWIFT Structured (What If)	+	+	-	+	+	1	+	+	+
Multiple-criteria decision analysis	+	+	+	+	+	3	-	+	-
Root Cause Analysis	-	+	+	-	+	1	+	+	+
Scenario Analysis	+	+	+	-	+	1	+	-	+
Decision Tree Analysis	-	+	+	+	+	2	+	+	-
Fault Tree Analysis	+	-	+	+	+	2	+	-	+
Event Tree Analysis	+	+	+	+	-	2	-	+	-
Cause/Consequence Analysis	+	+	-	+	+	1	+	+	+
Cause and Effect Analysis	+	+	-	-	-	1	+	-	+
FMEA and FMECA	+	+	+	+	+	3	+	+	+
Modified MEFD	+	-	-	+	+	3	+	+	+
Reliability Centered Maintenance	+	+	+	+	+	3	-	-	+
Consequence/probability matrix	-	+	+	+	+	2	+	+	-
HACCP (hazard analysis and critical control)	+	+	-	-	+	3	-	-	+
Bow Tie Analysis	-	+	-	-	-	2	+	-	+
Markov Chains Analysis	+	+	-	-	-	5	-	+	-
Monte Carlo Analysis	-	-	-	-	+	4	+	+	-
Bayesian Analysis	-	+	-	-	-	5	-	+	+

Table 2 Risk assessment techniques characterization

In order to make an accurate risk assessment is required to choice the appropriate method suitable for the industry. There are six groups of risk assessment techniques with examples of techniques below [10]:

- look-up methods (checklist, preliminary hazard analysis),
- supporting methods (brainstorming, delphi technique, SWIFT method, human reliability analysis),
- scenario analysis (root cause analysis, scenario analysis, fault tree analysis, event tree analysis, cause/consequence analysis, cause and effect diagram),
- functional analysis (FMEA and FMECA, reliability centered maintenance),
- controls assessments (LOPA-layers of protection analysis, bow tie analysis),
- statistical methods (markov analysis, monte carlo analysis, bayesian analysis [1, 12]).

The analysis of risk assessment can be qualitative, quantitative or a combination of these, depending on the circumstances. Depending on the type of the techniques and the difficulty in their implementation some methods were characterized in the context of production in Table 2. As a factors we chosen the type of method (quantitative, qualitative or both) and the level of difficulty to apply them in the company from 1 – 5, beginning from the easiest to the most difficult ones. The table also marks the information (+) whether the tool can be used in common practice and some elements of risk assessments process if they exist. They were marked for all tools in Table 2. There is also one technique that we propose - Modified MEFD (modified Matrix Errors Flow Diagram) as a tool to identify failures and provide reduction of risk in a set of techniques [15]. The proposed method and the proposed parameter to risk assessment provide a clear and consistent illustration of errors placement what can minimize potential problems in the future.

5 The proposal for a procedure

From the reason that none of method not allows to make a full risk assessment analysis we propose to use a mixed method based on the risk assessment techniques characterization (Table 2). The mixed method study technique can be applied by the following procedure:

Step I - use one or more of the methods which allows to identify all risks in the company. These are the methods marked (+) in a column in *risk detection.*

Step II - use one or more of the methods which allows to *analysis of risk* with indicators of consequence, probability and level of risk. More methods can be chosen when they are complemented each other.

Step III - use one of the methods which allows to prepare *risk evaluation*.

During the selection of appropriate technique take to account the level of difficulties of implementation and try to involve both qualitative and quantitative.

6 The case study

The area of this study refers to the moulding process. Moulding is a technology used to cast iron components to a specific design and shape. Moulding allows for the production of used objects by pouring molten metal into premanufactured patterns and the finished items are known as castings. Castings are a significant part of machine body. The knowledge of the characteristics of castings is essential from the suitability point of view. A large number of factors involved in the processes of casting have an effect on the production and characteristics of the individual casting. From that reason the casting process is very complicated and unstable. To make a full risk assessment we followed by procedure below:

Step I - *1. the modified Matrix Flow Diagram* to list the real areas/processes (Table 3) where occur problems in the company based on data from quality control (internal shortcomings, mistakes and complaints from customers).
2. *the brainstorming method* to find all possible hazards which can detect in that marked processes and also in analyzed layout document.

Process no.	Process name	Process no.	Process name
1.	Analysis inquiry (AI)	19.	Pouring
2.	Production planning (PP)	20.	Self control of metal in the ladle
3.	Preparation of documents (PD)	21.	Shacking out of mould
4.	Production (P):	22.	Shot blasting
5.	Charge materials control	23.	Manual settling
6.	Loading of charge materials / melting	24.	Visual control after settling
7.	Charge composition control	25.	Hardness test
8.	Melting of cast iron	26.	Final control
9.	Chemical composition control	27.	Periodical control
10.	Molten metal in the ladle	28.	Milling

11.	Control of composition after modification	29.	Control of dimensions
12.	Temperature control of pouring	30.	Drilling and threading
13.	Forming materials control	31.	Control of dimensions
14.	Preparation of green sands	32.	Packing
15.	Control parameters of green sands	33.	Final control according to the CP
16.	Making of moulds	34.	Analysis inquiry (AI)
17.	Control of compact of mould	35.	Production planning (PP)
18.	Mould assembly	36.	Customer complaint (CC)

Table 3 Detected processes in a step I of the mixed method

Step II - *the cause and effect diagram* to detect causes and sub causes of the possible risks.

During the analysis of cause and effect we determined probability and effect if the risk occurs. Effects were presented as a costs next to the probability for every cause in a diagram (Figure 2). Chart was prepare in R program language

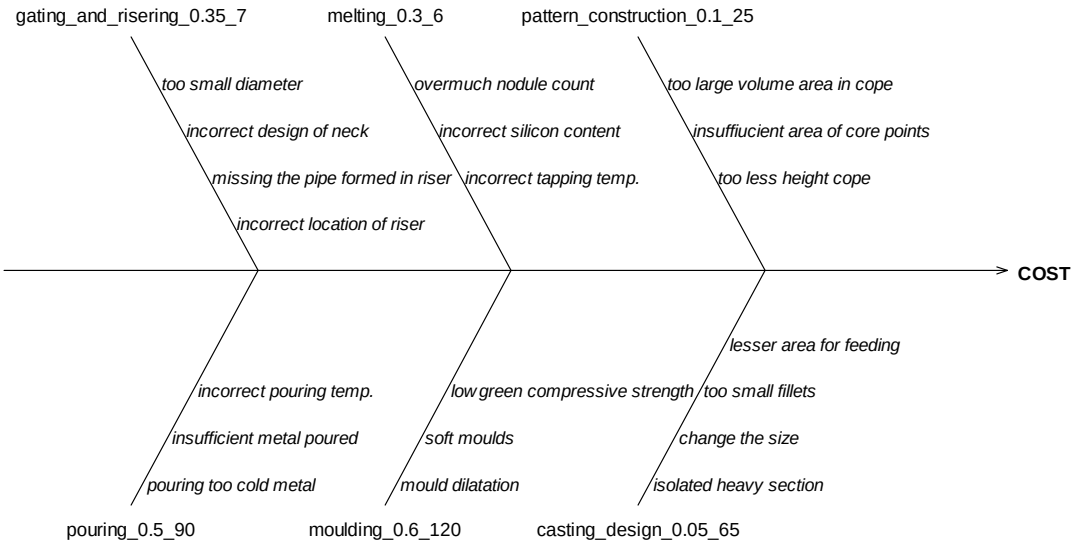


Figure 2 Cause and effect diagram for castings processes

Step III - *the monte carlo simulation* [11] method which takes into account the occurrence and the risk frequency. It allows to assess the impact of the identified risks on assumed goals, it requires knowledge of the risks. It requires knowledge of probability and the effect of risks expressed in numerical units. The effect of risk can be related to the aspects of the times of duration, quality and costs what is very needful in many companies.

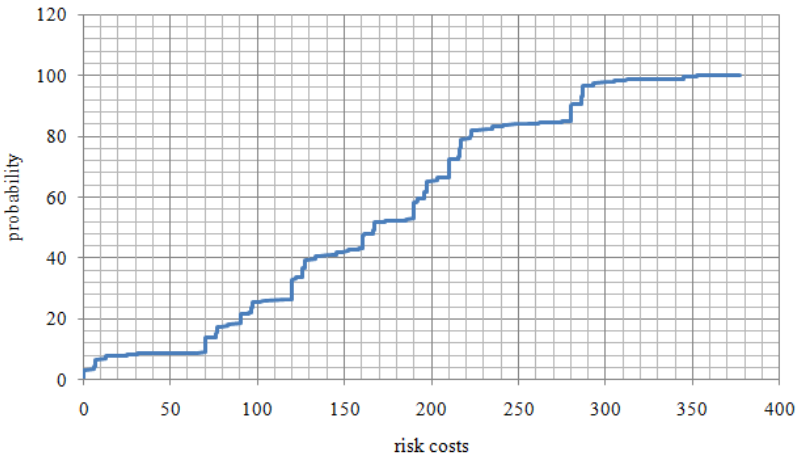


Figure 3 Cumulative distribution

In that method we analyzed impact of detected risks for goals of assumptions for producing castings. In that case impact of risks is related to the cost. The expected value of risk is the sum of the levels of individual risks. As a result of monte carlo simulation we present the cumulative distribution function which was perform for 1000 simulations (Figure 3).The most unfavorable scenario assumes that the risk will amount to 390 000 zloty. Assuming 80% confidence the risk assumes 220 000 zloty.

7 Conclusion

The quality risk management is an inherent part of all sectors of industry. Thus, all the statistical tools, mentioned in this paper can support and facilitate the risk management for quality. These tools can enable effective data estimation, assist in determining the significance of the data set, and to facilitate accurate decision-making.

The main objective of the analysis was to provide managers in organizations with some universal models and guidelines in the area of the quality risk management. It presented an idea of the mixed method study which consists of 3 steps procedure. It was illustrated step by step as a risk assessment tool.The proposed 3 steps method can be used by managers who are interested in monitoring the risks in processes without necessity of knowledge of other methods. There are some features which make this method attractive: simplicity, robustness, universality. By this method we can minimize time and cost, eliminate potential problem in the future and collect information about scope of all risks. It is a tool but more procedure which can be used with free software.

We noticed that is a need to construct one tool which would integrate some individual tools together. From that reason a proposal technique was introduced. We noticed the superiority of the proposed mixed method over other methods. This approach was proposed because integration of tools provides a better understanding of the risk assessment than either of each alone. It allows to make a full risk assessment analysis what cannot be done by a single technique for more industries.

There is lack of papers showing collected tools in the area of quality management that can be used in the risk assessment process. This article is an attempt to collect most of them. It can simplify the risk assessment process in quality management specially for production area.

References

[1] Aitkin, M., Boys, R.J., Chadwick, T.J. (2005). Bayesian point null hypothesis testing via the posterior likelihood ratio. *Statistics and Computing 15*, 217-230.

[2] Dai, F. C., Lee, C. F., Ngai, Y. Y. (2002). Landslide risk assessment and management: an overview, *Engineering Geology 64*, 65-87.

[3] Down, M. (2010). *Measurement System Analysis: reference manual*. Southfield, MI: Automotive Industry Action Group.

[4] FERMA (2002). *A Risk Management Standard*. Brussels.

[5] ISO (1991). ISO 8258. *Shewhart Control Charts.*

[6] ISO (1993a). ISO 7873. *Acceptance Control Charts.*

[7] ISO (1993b). ISO 7966. *Control Charts for Arithmetic Average With Warning Limits.*

[8] ISO (1997). ISO 7871. *Cumulative Sum Charts - Guidance on Quality Control and Data Analysis Using CUSUM Techniques.*

[9] ISO (2009a). ISO 73 Guide. *Risk Management - Vocabulary*.

[10] ISO (2009b). ISO 31000. *Risk Management - Principles and Guidelines.*

[11] ISO (2009c). ISO 31010. *Risk Management - Risk Assessment Techniques.*

[12] Li, Y., & You, J. (2012). Bayesian Hypothesis Testing in Latent Variable Models. *Journal of Econometrics 166*, 237-246.

[13] Montgomery, D. C., &Runger, G. C. (2003). *Applied statistics and probability for engineers*. New York: Wiley.

[14] Nomikos, P., & MacGregor, J. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics, 37*(1), 41-59.

[15] Rajda - Tasior, A. (2016). *The use of modified matrix errors flow diagram as a method of risk assessment for casting processes.* W: M Gliniak, J. Laskowska - Lemańska (red). Interdisciplinary issues of modern science - Monography (ss. 107 - 113). Kraków: Pro Futuro Science Association.

[16] Wawrzynek, J. (1999). *Metody statystyczne w zarządzaniu jakością.* W: W. Ostasiewicz (red.), Metody ilościowe w ekonometrii (ss. 337-373). Wrocław: Akademia Ekonomiczna.

Typology of consumers from Generation Y according to approach to the theatre

Katarína Rebrošová[1]

Abstract.
Theatre is one of the oldest art form and it has a long history. Theatre market has undergone considerable development and nowadays there are still groups of people, for whom is this kind of art very important. The aim of this work is to analyze the customer's attitude to the theatre and based on these attitudes create the customer's typology on the theatrical market in the Slovak Republic. Customers are from Generation Y.
For this purpose, it was used factor analysis followed by cluster analysis, which is in this article the main research method. It is used hierarchical cluster analysis for finding the number of clusters and even non-hierarchical cluster analysis, which assigned objects to concrete clusters. In the part of hierarchical clustering it is used Ward's procedure. Claims (attitudes) are defined on the basis of what kind of initiatives users show, when they are deciding about visiting theatre, what is their relationship to it, which types of performances they visit and what is the main point of visiting theatre.
Cluster analysis found 5 clusters (Visitors, who don't prefer theatre, Accompanying visitors, Occasional visitors, Theatre fans and Visitors following prestige) which showed big differences in their approach to the theatre.

Keywords: cluster analysis, factor analysis, Generation Y, theatre, typology.

JEL Classification: C38
AMS Classification: 62H30

1 Introduction

One of the people's choice, which is based also on the economic theory is division of time between the work and free time. People work, because they want to consume. Besides the fact, that people need to secure themselves and their families, save money for future, they spend money also on leisure activities.

At leisure activities there are also many different types of cultural performances including visiting theatre. Art faces a lot of challenges in the 21st century. These include the development of artificial markets at international level, the impact of economic restructuring on public financing, the change of definition of an art, as well as the interconnection of technologies and innovations.

This article deals with the theatrical market in the Slovak republic. Research is focused on the generation Y. It is a generation called "millennial" because it is marked by life of the people in two millennia. Hicks & Hicks [3] say, that individuals from this generation were born between 1980 and 2000. About this generation is told that its members don't like, when someone tells them, what to do and what not. They want to explore the world and build their own opinions.

The goal of this article is to analyse consumer's attitudes from the Generation Y on the theatrical market in Slovakia and then compile typology of consumers, which is based on these attitudes. There are tracked attitudes to the concrete types of performances (where act famous people, from the regional actors and performances based on some book), attitudes to the fact if a respondent is an initiator of visiting theatre or he is waiting for an invitation, how important is theatre for a respondent and if he/she really likes theatre).

2 Theoretical background

Which leisure activities person prefer is influenced by a number of factors such as the environment in which a person grows, the impact of reference groups, the personality of a person, or his / her financial capacity and physical predisposition. The research of leisure activities is based on the findings of article from Akar et al [1], which examined the impact of various demographic and social factors on the fact, which leisure activities consumers prefer. The cluster analysis method was used to identify relationships between attributes to activities.

Vysoká škola báňská – Technical university of Ostrava, Department of marketing and business, Sokolská třída 33, 702 00 Ostrava 1, katarina.rebrosova@vsb.cz

Ollivier, Gauthier, and Truong [5] examined the popularity of the various leisure activities. They focused on a large number of cultural activities which are nowadays known. It was for example visiting musicals, ballets, theatre, concerts, watching TV, reading books, magazines, listening to radio, visiting cinemas and other. These activities have been tracked in time, from 1979 to 2004. The result was that visiting theatre and its various genre branches (musicals, ballet) has been declining. We notice a significant increase in visiting cinema.

The need of a typology of theatre market customers is certainly considerable, but there are no papers which are describing theatre environment with its specific characteristics in connection with typology of consumers. This was the reason for doing it. Each institution needs to know its customers, so they can understand how to access them, what marketing tools can be used to capture them and communicate with them. So, Slack, Rowley and Coles [8] say that it is important to know the channel that can reach a particular customer and which has one or more of the following goals:

- deliver entertainment or information to audience to help with their purchasing decisions,
- engaging customers into marketing communication with the community,
- facilitate transactions related to the purchase of goods and services,
- deliver products or services to customers according to their preferences and needs,
- build relationships with customers and manage processes related to building and maintaining trust and loyalty.

3 Methodology

In this chapter are written procedures, which were used for data analysis and which lead to the achievement of the objectives of this article.

3.1 Data collecting

For the data collection that was needed to use cluster analysis, electronic querying was chosen. Electronic querying was selected mainly for suitability with regard to theme of an article, but also due to time and low cost. As it was mentioned in the introduction, the basic set was composed of respondents from Generation Y, so they were between 17 and 35 years old.

The respondents were selected on the basis of the quasi-representative technique, which was concretely the quota choice. Two quotas were used, namely sex of respondents and residence (city, village). Based on information from the Slovak statistical office an author of this article found the concrete proportion between the 17 till 35 years old respondents according to concrete quotas. During the querying it was not possible to abide these quotas, so author of article used data weighting in SPSS. The concrete data, which were used in weighting you can see in a *Table 1*.

		City %	Village %
Slovak population	Men %	27	24
	Women %	26	23
Slovak sample as a result of questioning	Man %	16	25
	Women %	26	33
Variables used for data weighting	Man	1.69	0.96
	Women	1	0.7

Table 1 Quotas used for data weighting

The basis of a research is on the attitude question, which were used in the diploma thesis, who's author is Rebrošová [6]. The question consists of 10 arguments. The task of all respondents was to express agreement with the concrete arguments. Likert scale was used with the values from 1 to 5, where 1 means „I fully agree" and 5 means „I totally disagree". Attitude question includes the following arguments:

- T1- I search theatre performances, where act famous actors.
- T2 - I search performances from the regional theatre actors.
- T3 - I like to visit performances, which are based on the books.
- T4 - Theatre is for me the most searched type of cultural action.
- T5 - I visit theatre performances only when someone invites me.
- T6 - I initiate the theatre visiting on my own.

- T7 - Visiting theatre is my hobby.
- T8 - By visiting theatre I express my social status.
- T9 - I perceive the theatre as a way of cultural development of my personality.
- T10 - I can't imagine my life without theatre.

Data has according the information by Řezánková [7] ordinal character and author works with the sample of 181 respondents. Response rate of a questionnaire was about 76 %. According to the theme of an article and the goal of it, author used cluster analysis as a main method. Before the cluster analysis it was also used factor analysis. Factor analysis was used to reduce factors for better interpretation in cluster analysis. Cluster analysis divided respondents into segments.

3.2 Methods of data analysis

Factor analysis is defined by Hebák [2] as a multidimensional statistical method that focuses on generating new variables and also data reduction to avoid losing information. In this case, factor analysis was used to identify the basic factors to identify smaller groups within a large dataset and to. Author used the exploratory factor analysis, because there don't exist typologies of consumers from the theatrical market, so there are no hypotheses that could be verified in the confirmative factor analysis. Author's plan is to be interested in confirmation factor analysis in the next researches.

Hebák [2] continues that one of the very important goals of factor analysis is to review the structure of relations between the tracked variables and get notice, if it is possible divide them into groups, in which variables from the same groups correlate more than variables from the different groups. The next goal of factor analysis is to create new uncorrelated variables – factors, which should better understand analysed data or another reasons to use.

The estimation of the factor analysis model is based on the principal component analysis. Hebák [2] and Řezánková [7] say that the basic approach is to reduce the size of a task and to create new (auxiliary) variables. If the variables Xj (j = 1,2, ..., m) are normalized (their mean is zero and the standard deviation is one), we can write them as it is in *Eguation 1*:

$$x_j = \sum_{l=1}^{L} \lambda_{jl} F_l + \varepsilon_j, \tag{1}$$

where Xj are variables, which we want to express,
j means order of a variable,
F_l means L common factors, which cause correlations between variables,
E_j are specific (false) factors, which contribute to the variance of each of the observed variables,
λ_{jl} are weights, which are called also factor loadings.

To find out the appropriateness of using factor analysis as a tool for a reduction of arguments was used Kaiser-Maier-Olkin test (KMO) and Bartlett's test of Sphéricity. KMO value is 0.833 and it means that it is possible to use factor analysis, because a boundary value for it as a tool for factors reduction it should be higher than 0,5. We can find in the various literature that if value is closer to 1 it is better for application of factor analysis. Bartlett's test of Sphéricity was the second test. The null hypothesis at the Bartlett's test of Sphéricity is formulated as a no correlation between the tested variables. Significance value is less than 5%, so we denied H_0 and between the tested variables is correlation. We can really use factor analysis. The results of both test you can see in a *Table 2*.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		**.833**
Bartlett's Test of Sphericity	Approx. Chi-Square	570.511
	Df	45
	Sig.	**.000**

Table 2 The result of appropriateness of using factor analysis

In the next step author wanted to reduce the number of variables and find out the number of new factors. For this purpose, Kaiser rule was used. It means that are used only those factors, whose own value is greater than one

(Eigenvalue value is greater than one). For factor rotation was used an orthogonal Varimax method and factor scores are obtained on the basis of regression method. Eigenvalue is higher than one for the case of two new factors. In the case of reduction of ten arguments into two new factors is the variance explained almost from 54 %.

Two new factors have been created, which can be seen in *Table 3* also with the fact, which arguments they include. The relevance of the arguments to the new factors is determined by their values in these factors. The argument is included in that factor in which it has a higher value, regardless of whether the value is positive or negative. Factors have been named according aspects that lead to a visiting theatre. It can be seen that the first factor contains six arguments and on the basis of their content, it is possible to call this factor a "psychological aspect". It means that people go to the theatre, because of their deep motivation and because they are interested in it. In contrast, the second factor contains only four arguments, but it follows a certain speculative behaviour of consumers that leads to choosing theatrical performances because they want to become visible in the society and build their social status. Therefore, the second factor is called the "sociological aspect".

	Component	
	Psychological aspect	Sociological aspect
T7	0.820	0.191
T10	0.807	0.153
T6	0.792	
T4	0.754	0.210
T9	0.679	0.155
T5	-0.595	0.381
T1	-0.297	0.663
T2	0.205	0.647
T3	0.167	0.600
T8	0.464	0.512

Table 3 Component rotated matrix

Author of article continued with the cluster analysis. She also tried to use it with the original ten arguments. Results were very similar, but there is recommendation to use reduced data from factor analysis as an enter into cluster analysis because of better interpretation and also better possibility to find out differences between clusters.

4 Creation of consumer's typology

For the result of this article, the most suitable was using hierarchical and also non-hierarchical clustering. Hierarchical clustering was used to find the concrete number of clusters. Here it was used Ward procedure and as a distance, author of this article used squared Euclidean distance, which is as Malhotra mentions [4] one of the most used.

At first it was necessary to find the number of clusters, with which is possible to work in the following steps. Into hierarchical clustering entered factor loadings, which were created in factor analysis. As it was mentioned, it was used Ward procedure and squared Euclidean distance. Řezánková [7] describes that Ward procedure says that are joining clusters, in which is increase of the total intra-group summary of squares of deviations of individual values from cluster mean minimal. This method proved to be one of the best and its formula, which is in *Eguation 2* is:

$$D_{g<h,h'>} = \frac{(n_h + n_g)D_{gh} + (n_{h'} + n_g)D_{gh'} - n_g D_{hh'}}{n_h + n_{h'} + n_g}, \tag{2}$$

where D means a distance between object g and unification of objects h and h',

About final number of clusters decided the difference of the coefficients. Useful was also dendrogram. Because the decision about the number of clusters is often quite subjective and depends mostly on the researcher, the author of the work contemplates to use four, five or six clusters. In the final decision about their number, author relied on the average values of the arguments and the ANOVA test, where the level of significance was monitored. Required value was less than 0.05. Finally, five clusters / segments were created.

The research continued with the k-means method, which belongs into non-hierarchical methods of clustering. This method divided respondents into the following five segments:

- 1. Segment – Visitors, who don't prefer theatre - 25 %
- 2. Segment – Accompanying visitors – 28.5 %
- 3. Segment – Occasional visitors – 18.9 %
- 4. Segment – Theatre fans – 11.2 %
- 5. Segment – Visitors following prestige – 16.4 %

Clusters got names according the average values of the concrete arguments, which you can see in a *Table 4*. Character of clusters is derived from the information in the table. As it was mentioned, values of arguments are from 1 to 5. Number one means the total agreement with the argument and 5 means total disagreement with it.

The first group of consumers can be called " visitors, who don't prefer theatre ". These consumers don't show sympathy for the theatrical environment, they don't perceive it as something essential and useful for their lives. They do not initiate visiting theatre, so it can be assumed that this group visit theatre only as some kind of duty (work, social, personal).

	Visitors, who don't prefer theatre	Accompanying visitors	Occasional visitors	Theatre fans	Visitors following prestige
I search theatre performances, where acts famous actors.	2.441	2.573	3.131	3.712	1.794
I search performances from the regional theatre actors.	3.573	3.077	3.126	3.387	1.856
I like to visit performances, which are based on the books.	3.239	2.36	3.071	2.492	2.065
Theatre is for me the most searched type of cultural action.	4.58	3.946	3.879	2.001	2.597
I visit theatre performances only when someone invite me.	2.385	1.807	4.06	4.442	3.502
I initiate the theatre visiting myself.	4.678	4.244	2.78	1.804	2.686
Visiting theatre is my hobby.	4.563	2.976	3.312	1.256	2.15
By visiting the theatre I express my social status.	4.437	3.034	3.962	3.051	2.316
I perceive the theatre as a way of cultural development of my personality.	3.729	2.202	1.996	1.422	1.99
I can't imagine my life without theatre.	4.93	4.105	4.272	1.624	2.434

Table 4 Average values of arguments in the concrete clusters

The second group of customers consists of "accompanying visitors". As the name itself suggests, they are individuals who also do not search usually theatre performances. It is even typical for them to visit theatre only when someone invites them. So they are accompanying someone. They don't have such a negative relationship to the theatre as the first group of consumers and on the contrary, they consider it to be useful for the development of a human personality.

The next group of customers is named "occasional visitors". There are consumers who are socially active and visit various cultural performances. Theatre is not for them the first choice, but they don't have a negative attitude to it, and from time to time they also visit theatre performances, whether as initiators or they are invited. They also think that theatre has a positive influence to people.

The fourth group is called "theatre fans". This group includes consumers, who take the theatre as an irreplaceable part of their lives. It belongs to their hobbies and also see significant benefits for their personality. They don't wait for invitations from someone (family, friends, etc.), but on the contrary they initiate a visit of performances according to their preferences. Of all types of cultural events, the theatre is the one that they visit most often.

The last group is called "visitors following prestige". Based on the average of the individual arguments, this group of consumers is closely monitoring individual types of performances. From a wide range of theatrical performances, they prefer those, where are famous actors, plays from the same region as consumer, or are based on a

book story. For this group is typical that it is attending theatre because of prestige and in order to make more visible in society and increase the own social status.

5 Conclusion

The aim of this article was to analyse the attitudes of customers on the theatrical market in the Slovak Republic. The respondents were from generation Y. The theoretical background of the studied area was described and it continued with the methodological part of the research. To achieve these aims, cluster analysis was selected as the main method.

The author used cluster analysis to solve the problem. In the initial phase, Ward's procedure was used as one of the hierarchical methods to determine the number of clusters. The result from hierarchical clustering was used in the following non-hierarchical method, which was the k-means method. It divided the individual respondents into a predetermined number of clusters. Typology showed 5 segments on the Slovak theatrical market.

A relatively large group of people visit the theatre, even they don't like it and they don't prefer this kind of culture. It is also possible to find on the theatrical market those, who occasionally visit the theatre. There are two segments of these people. The first one consist of people, who are accompanying somebody and in the second segment are occasional visitors, who think that theatre can influence their life and personality in a good way. In the author's view, theatres could be more focused on a segment of people, who are not a big fans of theatre, but visit it from time to time (accompanying and occasional visitors) because they are quite a large group.

The specific group is “theatre fans”. They are people who really enjoy the theatre and it is an essential part of their leisure activities. The last group consists of visitors following the prestige and it is important for them to be seen in the theatres, because they want to build a good name in the society.

References

[1] Akar, G., Clifton K. J., and Doherty, S. T.: Redefining activity types: Who participates in which leisure activity? In: *Transportation Research Part A: Policy and Practice* **46** (2012), 1194-1204.
[2] Hebák, P., Hustopecký, J., Pecáková, I., Plašil, M., Průša, M., Řezanková, H., Svobodová, A., and Vlach, P.: *Vícerozměrné statistické metody (3)*. Informatorium, Praha, 2007.
[3] Hicks, R., avnd Hicks, K.: Boomers, X-ers, and Other Strangers: Understanding/Generational Differences/Divide Us. Focus on the Family, 1999.
[4] Malhotra, N. K.: *Marketing Research. An Applied Orientation.* Prentice Hall, New Jersey, 2010.
[5] Ollivier, M., Gauthier, G., and Truong, A. H.: Cultural classifications and social divisions: A symmetrical approach. In: *Poetics* **37** (2009), 456-473.
[6] Rebrošová, K.: Typologie chování spotřebitelů na divadelním trhu v mezinárodním kontextu. Diploma thesis, Ostrava, 2016.
[7] Řezánková, H., Húsek D., and Snášel, V.: *Shluková analýza dat (druhé rozšířené vydání)*. Professional Publishing, Praha, 2009.
[8] Slack, F., Rowley, J., and Coles, S.: Consumer behaviour in multi channel contexts: The case of a theatre festival. *Internet Research* **18** (2008), 46–59.

Decision matrices under risk with fuzzy states of the world and underlying discrete fuzzy probability measure

Pavla Rotterová[1], Ondřej Pavlačka[2]

Abstract. The problem of decision-making under risk can be expressed by a decision matrix whose elements express the outcomes if a decision-maker chooses the particular alternative and the particular state of the world occurs. In practice, the states of the world can be described only vaguely. Therefore, in such a case we dealt with the decision matrix with fuzzy states of the world that were modelled by fuzzy sets on the universal set on which the probability distribution is given. However, the underlying probability distribution itself can be known also only vaguely. In such a case it is appropriate to apply a fuzzy probability measure.
In the paper, we consider the case where the set of elementary events is finite and the probabilities of elementary events are fuzzy. We compare two approaches to expression of the expected values and the variances of the outcomes of alternatives. The first approach is inspired by the probability of a fuzzy event proposed by Zadeh. The second one is based on treating the decision matrix as a fuzzy rule bases system. We illustrate the problem by a numerical example from economic practice.

Keywords: decision matrices, decision-making under risk, fuzzy probability measure, fuzzy states of the world, fuzzy rule-based systems.

JEL classification: C44
AMS classification: 90B50

1 Introduction

A decision matrix, usually given as a table, is often used as a tool of risk analysis (see e.g. [1, 9]). It describes how outputs of alternatives $x_1, \ldots, x_n$ depend on the fact which of possible and mutually disjoint states of the world $S_1, \ldots, S_m$ will occur. The probabilities of occurrences of the states of the world, denoted by $p_1, \ldots, p_m$, are assumed to be known. Thus, the outcome of choosing an alternative x_i, $i \in \{1, \ldots, n\}$, is a discrete random variable H_i that takes on the values $h_{i,1}, \ldots, h_{i,m}$ with probabilities $p_1, \ldots, p_m$. The alternatives are usually compared on the basis of the expected values $EH_1, \ldots, EH_n$ and variances $varH_1, \ldots, varH_n$ of their outcomes.

In practical applications, the states of the world are often described only vaguely, like e.g. "grow rate of gross domestic product is high", etc. Talašová and Pavlačka [8] showed that in such a case it is more appropriate to model the states of the world by fuzzy sets on the universal set on which the probability distribution is given. They proposed to proceed in the same way as in the case of exactly described, i.e. so-called crisp, states of the world. They set the probabilities of the fuzzy states of the world applying the formula proposed by Zadeh [10] and treated the outputs of alternatives as discrete random variables.

However, in [3] and [6] it was shown that Zadeh's probabilities of fuzzy events lack the common interpretation of a probability measure. Another problem represents a meaning of a statement that the particular state of the world will occur. Therefore, in [4] and [5] it was proposed another approach to the considered problem where the decision matrix is understood as a system of rule bases. Moreover, in economical applications the probability measure of the universe, on which fuzzy states of the world are defined, can be fuzzy, e.g. set by experts. Thus, in this paper we will consider a discrete fuzzy probability measure (see Section 2). Then, we will extend the approach using Zadeh's probabilities to the case of fuzzy probabilities. In Section 3, we will describe the considered problem by a fuzzy rule bases system. Finally Conclusion finishes the paper.

2 A decision matrix with fuzzy states of the world and with fuzzy probabilities of these states of the world

First, let us describe the extension of a decision matrix to the case of fuzzy states of the world where probabilities of the elementary events are expressed by so-called fuzzy numbers.

[1]Palacký University Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic, pavla.rotterova01@upol.cz
[2]Palacký University Olomouc, Faculty of Science, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 771 46 Olomouc, Czech Republic, ondrej.pavlacka@upol.cz

Let us assume that a probability space $(\Omega, \mathcal{A}, P)$ is given, where $\Omega = \{\omega_1, \ldots, \omega_r\}$ denotes a non-empty set of all elementary events, $\mathcal{A}$ represents the set of all considered random events ($\mathcal{A}$ forms a σ-algebra of subsets of Ω), and $P : \mathcal{A} \to [0, 1]$ is a probability measure that assigns to each random event $A \in \mathcal{A}$ its probability $P(A) \in [0, 1]$.

A fuzzy state of the world can be appropriately expressed by a fuzzy set S on Ω. A *fuzzy set S on Ω* is determined by its membership function $\mu_S : \Omega \to [0, 1]$. The interpretation of membership degrees $\mu_S(\omega)$, $\omega \in \Omega$, follows: Let us consider a fuzzy set "high loss". Then, its membership degrees are usually interpreted as follows: $\mu_S(\omega) = 1$ means that the achieved loss is definitely high. $\mu_S(\omega) = 0$ means that the loss is not high. $\mu_S(\omega) \in (0, 1)$ means that the loss is high only partly.

As the probability space $(\Omega, \mathcal{A}, P)$ is given, in [8] it was assumed that fuzzy states of the world are expressed by fuzzy sets on Ω that are called fuzzy random events. A *fuzzy random event S* is a fuzzy set on Ω whose membership function μ_S is $\mathcal{A}$-measurable (see Zadeh [10]). This assumption means that the α-cuts $S_\alpha := \{\omega \in \Omega \mid \mu_S(\omega) \geq \alpha\}$, $\alpha \in (0, 1]$, are random events, i.e. $S_\alpha \in \mathcal{A}$ for any $\alpha \in (0, 1]$.

For the case of fuzzy random events, Zadeh [10] extended the given probability measure P in the following way: A probability $P_Z(S)$ of a fuzzy random event S is defined as

$$P_Z(S) := E(\mu_S) = \int_\Omega \mu_S(\omega) dP. \tag{1}$$

It can be easily shown (see e.g. [3, 10]) that the mapping P_Z possesses analogous mathematical properties as a probability measure P, and thus, it can be called a probability measure.

Remark 1. Let us note that any crisp set $S \subseteq \Omega$ can be viewed as a fuzzy set of a special kind; the membership function μ_S coincides in such a case with the characteristic function χ_S of S. In such a case, $S_\alpha = S$ for all $\alpha \in (0, 1]$. This convention allows us to consider crisp states of the world as a special kind of fuzzy states of the world, for which $P_Z(S) = P(S)$ for any crisp random event $S \in \mathcal{A}$.

Now, let us extend the probability $P_Z(S)$ of the fuzzy state of the world S to the case of a fuzzy probability $P_\mathcal{F}(S)$ of the state of the world S, which is expressed by a fuzzy number.

A *fuzzy number S* is a fuzzy set on $\mathbb{R}$ such that its core $Core\ S := \{\omega \in \Omega \mid \mu_S(\omega) = 1\}$ is non-empty, its α-cuts S_α are closed intervals for any $\alpha \in (0, 1]$, and its support $Supp\ S := \{\omega \in \Omega \mid \mu_S(\omega) > 0\}$ is bounded. The family of all fuzzy numbers, whose supports are subsets of [0,1], is denoted by $\mathcal{F}_N([0, 1])$.

Let it holds that $P_\mathcal{F}(\{\omega_k\}) \in \mathcal{F}_N([0, 1])$, $k \in \{1, \ldots, r\}$. Let $P_\mathcal{F}(\{\omega_1\}), \ldots, P_\mathcal{F}(\{\omega_r\})$ form an *r-tuple of fuzzy probabilities* if for any $j \in \{1, \ldots, r\}$ and any $\alpha \in (0, 1]$ it holds that for any $p_j \in P_\mathcal{F}(\{\omega_j\})_\alpha$ there exist $p_k \in P_\mathcal{F}(\{\omega_k\})_\alpha, k = 1, \ldots, r, k \neq j$, such that $p_j + \sum_{k=1, k\neq j}^{r} p_k = 1$. A *fuzzy probability $P_\mathcal{F}(S)$ of the fuzzy state of the world S* is computed by its α-cuts. For any $\alpha \in (0, 1]$, $P_\mathcal{F}(S)_\alpha = [p^L_{S\alpha}, p^U_{S\alpha}]$ is computed as follows:

$$p^L_{S\alpha} = \min \left\{ \sum_{k=1}^{r} \mu_S(\omega_k) \cdot p_k \mid p_k \in P_\mathcal{F}(\{\omega_k\})_\alpha, k = 1, \ldots, r, \sum_{k=1}^{r} p_k = 1 \right\}, \tag{2}$$

and

$$p^U_{S\alpha} = \max \left\{ \sum_{k=1}^{r} \mu_S(\omega_k) \cdot p_k \mid p_k \in P_\mathcal{F}(\{\omega_k\})_\alpha, k = 1, \ldots, r, \sum_{k=1}^{r} p_k = 1 \right\}. \tag{3}$$

Obviously, the formulas (2) and (3) represent fuzzy extensions of the formula (1) for a discrete finite universal set Ω. Thus $P_\mathcal{F}$ is a *discrete fuzzy probability measure* (fulfils boundary conditions, i.e. $P_\mathcal{F}(\Omega) = 1$ and $P_\mathcal{F}(\emptyset) = 0$, and monotonicity, i.e. $P_\mathcal{F}(A) \leq P_\mathcal{F}(B)$ for any $A \subseteq B$ (see [2])).

At the end of the section, let us describe the fuzzy extension of the decision matrix mentioned in Introduction. The fuzzy states of the world are fuzzy random events $S_1, \ldots, S_m$ that form a *fuzzy partition of Ω*, i.e. $\sum_{j=1}^{m} \mu_{S_j}(\omega) = 1$ for any $\omega \in \Omega$. The probabilities $P_{\mathcal{F}1}, \ldots, P_{\mathcal{F}m}$ of the fuzzy states of the world are given by $P_{\mathcal{F}j} := P_\mathcal{F}(S_j)$, $j = 1, \ldots, m$. Since $S_1, \ldots, S_m$ form a fuzzy partition of Ω, $P_{\mathcal{F}1}, \ldots, P_{\mathcal{F}m}$ form the m-tuple of fuzzy probabilities. Therefore, the output, if an alternative x_i, $i \in \{1, \ldots, n\}$, is chosen, can be treated as a discrete random variable H^Z_i that takes on the values $h_{i,1}, \ldots, h_{i,m}$ with fuzzy probabilities $P_{\mathcal{F}1}, \ldots, P_{\mathcal{F}m}$. Let us compare the alternatives $x_1, \ldots, x_n$ on the basis of the fuzzy expected values $EH^Z_1, \ldots, EH^Z_n$ and fuzzy variances $var\, H^Z_1, \ldots, var\, H^Z_n$ of their outputs. For $i = 1, \ldots, n$ and for any $\alpha \in (0, 1]$, the α-cut $EH^Z_{i,\alpha} = \left[Eh^{ZL}_{i,\alpha}, Eh^{ZU}_{i,\alpha}\right]$ of the fuzzy expected output is obtained as follows:

$$Eh^{ZL}_{i,\alpha} = \min \left\{ \sum_{j=1}^{m} p_j \cdot h_{i,j} \mid p_j \in (P_{\mathcal{F}j})_\alpha, j = 1, \ldots, m, \sum_{j=1}^{m} p_j = 1 \right\} \tag{4}$$

	S_1	S_2	$\cdots$	S_m		
	$P_{\mathcal{F}1}$	$P_{\mathcal{F}2}$	$\cdots$	$P_{\mathcal{F}m}$		
x_1	$h_{1,1}$	$h_{1,2}$	$\cdots$	$h_{1,m}$	EH_1^Z	$var\ H_1^Z$
x_2	$h_{2,1}$	$h_{2,2}$	$\cdots$	$h_{2,m}$	EH_2^Z	$var\ H_2^Z$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_n	$h_{n,1}$	$h_{n,2}$	$\cdots$	$h_{n,m}$	EH_n^Z	$var\ H_n^Z$

Table 1 Considered Fuzzy Decision Matrix

and

$$Eh_{i,\alpha}^{ZU} = \max\left\{\sum_{j=1}^{m} p_j \cdot h_{i,j} \mid p_j \in (P_{\mathcal{F}j})_\alpha, j = 1,\ldots,m, \sum_{j=1}^{m} p_j = 1\right\}. \tag{5}$$

The α-cut $var\ H_{i,\alpha}^Z = \left[var\ h_{i,\alpha}^L, var\ h_{i,\alpha}^U\right]$ is obtained as follows: Let us denote

$$z_i(p_1,\ldots,p_m) = \sum_{j=1}^{m} p_j \cdot \left(h_{i,j} - \sum_{t=1}^{m} p_t \cdot h_{i,t}\right)^2.$$

Then,

$$var\ h_{i,\alpha}^L = \min\left\{z_i(p_1,\ldots,p_m) \mid p_j \in (P_{\mathcal{F}j})_\alpha, j = 1,\ldots,m, \sum_{j=1}^{m} p_j = 1\right\} \tag{6}$$

and

$$var\ h_{i,\alpha}^U = \max\left\{z_i(p_1,\ldots,p_m) \mid p_j \in (P_{\mathcal{F}j})_\alpha, j = 1,\ldots,m, \sum_{j=1}^{m} p_j = 1\right\}. \tag{7}$$

Since just described approach is an extension of the approach for the fuzzy decision matrix with fuzzy states of the world and crisp Zadeh's probabilities (computed according to (1)) described in [4], the problems related with it are preserved. The first problem is that Zadeh's probabilities of fuzzy events lack the common interpretation of a probability measure. Another problem is a precise definition of "the occurrence of the particular fuzzy state of the world". A more detailed description of this problems can be found in [4, 6]. Therefore, interpretations of the fuzzy expected outputs $EH_1^Z,\ldots,EH_n^Z$, given by (4) and (5), and $var\ H_1^Z,\ldots,var\ H_n^Z$, given by (6) and (7), is difficult. Thus, ordering of the alternatives based on these characteristics is problematic.

3 Fuzzy rule-based systems derived from the considered decision matrix

In this section, we introduce a different approach to the decision matrix with fuzzy states of the world. Taking into account the problems mentioned in the previous section, we suggest not to treat the output, if an alternative x_i, $i \in \{1,\ldots,n\}$, is chosen, as a discrete random variable H_i^Z taking on the values $h_{i,1},\ldots,h_{i,m}$ with the fuzzy probabilities $P_{\mathcal{F}1},\ldots,P_{\mathcal{F}m}$. Instead of this, we propose to model the information about the output of choosing x_i as the following basis of If-Then rules:

$$\begin{aligned}&\text{If the state of the world is } S_1, \text{ then the consequence of } x_i \text{ is } h_{i,1}.\\ &\text{If the state of the world is } S_2, \text{ then the consequence of } x_i \text{ is } h_{i,2}.\\ &\ldots\\ &\text{If the state of the world is } S_m, \text{ then the consequence of } x_i \text{ is } h_{i,m}.\end{aligned} \tag{8}$$

For computing the output of the fuzzy rule-based system (8), it is appropriate to apply the so-called *Sugeno method of fuzzy inference*, introduced by Sugeno [7]. For any $\omega_k \in \Omega$, $k \in \{1,\ldots,r\}$, the consequence of choosing x_i is given by

$$H_i^S(\omega_k) = \frac{\sum_{j=1}^{m} \mu_{S_j}(\omega_k) \cdot h_{i,j}}{\sum_{j=1}^{m} \mu_{S_j}(\omega_k)} = \sum_{j=1}^{m} \mu_{S_j}(\omega_k) \cdot h_{i,j}, \tag{9}$$

where we used the assumption $\sum_{j=1}^{m} \mu_{S_j}(\omega_k) = 1$ for any $\omega_k \in \Omega$. Since we operate within the given probability space $(\Omega, \mathcal{A}, P)$, H_i^S is a random variable.

Remark 2. It can be easily seen from (9) that in the case of crisp states of the world $S_1, \ldots, S_m$, the random variables $H_1^S, \ldots, H_n^S$ coincide with discrete random variables $H_1, \ldots, H_n$ taking on the values $h_{i,1}, \ldots, h_{i,m}$, $i = 1, \ldots, n$, with fuzzy probabilities $P_{\mathcal{F}j} = P_{\mathcal{F}}(S_j)$, $j = 1, \ldots, m$. Hence, this new approach can be also considered as an extension of a decision matrix to the case of fuzzy states of the world and their fuzzy probabilities.

Analogously as in the previous approach, the alternatives $x_1, \ldots, x_n$ can be compared on the basis of the expected values and variances of $H_1^S, \ldots, H_n^S$. Let us derive now the formulas for their computation. In addition to that, we will compare these characteristics with the characteristics of $H_1^Z, \ldots, H_n^Z$.

First, let us show that the expected values of $H_1^S, \ldots, H_n^S$ coincide with $EH_1^Z, \ldots, EH_n^Z$. For $i = 1, \ldots, n$, and any $\alpha \in (0, 1]$, let us denote $EH_{i,\alpha}^S = \left[Eh_{i,\alpha}^{SL}, Eh_{i,\alpha}^{SU}\right]$. We get

$$\begin{aligned} Eh_{i,\alpha}^{SL} &= \min\left\{\sum_{k=1}^{r} H_i^S(\omega_k)\cdot p_k \mid p_k \in P_{\mathcal{F}}(\{\omega_k\})_\alpha, k = 1, 2, \ldots, r, \sum_{k=1}^{r} p_k = 1\right\} \\ &= \min\left\{\sum_{k=1}^{r}\sum_{j=1}^{m} \mu_{S_j}(\omega_k)\cdot h_{i,j}\cdot p_k \mid p_k \in P_{\mathcal{F}}(\{\omega_k\})_\alpha, k = 1, 2, \ldots, r, \sum_{k=1}^{r} p_k = 1\right\} \\ &= \min\left\{\sum_{j=1}^{m}\sum_{k=1}^{r} \mu_{S_j}(\omega_k)\cdot p_k\cdot h_{i,j} \mid p_k \in P_{\mathcal{F}}(\{\omega_k\})_\alpha, k = 1, 2, \ldots, r, \sum_{k=1}^{r} p_k = 1\right\} \\ &= \min\left\{\sum_{j=1}^{m} p_j\cdot h_{i,j} \mid p_j \in (P_{\mathcal{F}j})_\alpha, j = 1, \ldots, m, \sum_{j=1}^{m} p_j = 1\right\} = Eh_{i,\alpha}^{ZL} \end{aligned} \tag{10}$$

and analogously

$$\begin{aligned} Eh_{i,\alpha}^{SU} &= \max\left\{\sum_{k=1}^{r} H_i^S(\omega_k)\cdot p_k \mid p_k \in P_{\mathcal{F}}(\{\omega_k\})_\alpha, k = 1, 2, \ldots, r, \sum_{k=1}^{r} p_k = 1\right\} \\ &= \max\left\{\sum_{j=1}^{m} p_j\cdot h_{i,j} \mid p_j \in (P_{\mathcal{F}j})_\alpha, j = 1, \ldots, m, \sum_{j=1}^{m} p_j = 1\right\} = Eh_{i,\alpha}^{ZU}. \end{aligned} \tag{11}$$

Hence, from the point of view of the expected values of outputs that are taken into account in comparison of alternatives, both the approaches are the same.

Now, let us turn our mind to the variances of $H_1^S, \ldots, H_n^S$. The α-cut $var\, H_{i,\alpha}^S = \left[var\, h_{i,\alpha}^{SL}, var\, h_{i,\alpha}^{SU}\right]$ of the fuzzy variance of the output from the fuzzy rule-based system is obtained as follows: Let us denote

$$s_i(p_1, \ldots, p_r) = \sum_{k=1}^{r} p_k \cdot \left(\sum_{j=1}^{m} \mu_{S_j}(\omega_k)\cdot h_{i,j} - \sum_{t=1}^{r} p_t \cdot \sum_{u=1}^{m} \mu_{S_u}(\omega_t)\cdot h_{i,u}\right)^2.$$

Then,

$$var\, h_{i,\alpha}^{SL} = \min\left\{s_i(p_1, \ldots, p_r) \mid p_k \in P_{\mathcal{F}}(\{\omega_k\})_\alpha, k = 1, \ldots, r, \sum_{k=1}^{r} p_k = 1\right\} \tag{12}$$

and

$$var\, h_{i,\alpha}^{U} = \max\left\{s_i(p_1, \ldots, p_r) \mid p_k \in P_{\mathcal{F}}(\{\omega_k\})_\alpha, k = 1, \ldots, r, \sum_{k=1}^{r} p_k = 1\right\}. \tag{13}$$

Nevertheless, as we will show in the following example the variances of $H_1^S, \ldots, H_n^S$ are different from $var\, H_1^Z, \ldots, var\, H_n^Z$. This means that ordering of the alternatives, based on the expected values and variances of their outputs, can be different for both the above described approaches. We will illustrate this fact by the following numerical example.

Example 1. Let us consider that we can realize one of the two possible projects, denoted by A and B. The result depends solely on the kind of a government coalition which will be established after the parliamentary

election. Let us assume that only the six possible coalitions, denoted by $\omega_1, \ldots, \omega_6$, can be established after the election. For the sake of simplicity, let the fuzzy probabilities of establishing each coalition be the same equalled to $P_{\mathcal{F}}(\{\omega_k\}) = \langle 1/12, 1/6, 3/12\rangle$, $k = 1, \ldots, 6$, where three values in brackets represent *significant values* (boundary values of the closure of the support and the value of the core of the fuzzy number) of a *triangular fuzzy number* (the fuzzy number with linear membership functions between its significant values). $P_{\mathcal{F}}(S)$ is obtained by formulas (2) and (3), for any $S \in \mathcal{F}(\Omega)$, where $\mathcal{F}(\Omega)$ denotes the set of all fuzzy events whose support belongs to Ω.

We distinguish three possibilities (states of the world) - a right coalition (S_1), a centre coalition (S_2), and a left coalition (S_3). Let the above mentioned three states of the world be expressed by the following fuzzy sets defined on Ω:

$$\begin{aligned} S_1 &= \left\{ {}^{1}|_{\omega_1}, {}^{0.8}|_{\omega_2}, {}^{0.2}|_{\omega_3}, {}^{0}|_{\omega_4}, {}^{0}|_{\omega_5}, {}^{0}|_{\omega_6} \right\}, \\ S_2 &= \left\{ {}^{0}|_{\omega_1}, {}^{0.2}|_{\omega_2}, {}^{0.8}|_{\omega_3}, {}^{1}|_{\omega_4}, {}^{0.5}|_{\omega_5}, {}^{0}|_{\omega_6} \right\}, \\ S_3 &= \left\{ {}^{0}|_{\omega_1}, {}^{0}|_{\omega_2}, {}^{0}|_{\omega_3}, {}^{0}|_{\omega_4}, {}^{0.5}|_{\omega_5}, {}^{1}|_{\omega_6} \right\}, \end{aligned}$$

where elements of the sets are in the form ${}^{\mu_{S_j}(\omega_k)}|_{\omega_k}$, $j = 1, 2, 3$, and $k = 1, \ldots, 6$. How the future yield (in %) depends on the fact which of these three types of a coalition will occur in the future is described in Table 2.

	S_1	S_2	S_3
A	1	0.5	0.1
B	1	0.2	0.8

Table 2 Significant Values of Future Yields (in %) of A and B based on the Type of a Coalition

Now, let us compare both the approaches to extension of the decision matrix tool. First, let us compute the expected values and variances of the random variables H_A^Z and H_B^Z. The fuzzy probabilities of the fuzzy states of the world S_1, S_2, and S_3, computed according to (2) and (3), are given as triangular fuzzy numbers, whose significant values are given as follows:

$$P_{\mathcal{F}}(S_1) = \langle 0.17, 0.33, 0.50\rangle,\ P_{\mathcal{F}}(S_2) = \langle 0.21, 0.42, 0.63\rangle, \text{ and } P_{\mathcal{F}}(S_3) = \langle 0.12, 0.25, 0.37\rangle.$$

Hence, according to formulas (4) and (5) we get the fuzzy expected yields EH_A^Z and EH_B^Z expressed by triangular fuzzy numbers. Their significant values are shown in Table 3. The fuzzy variances $var\ H_A^Z$ and $var\ H_B^Z$, obtained by formulas (6) and (7), are not triangular fuzzy numbers. Their significant values are also shown in Table 3.

Second, let us construct the random variables H_A^S and H_B^S. Since the universal set Ω is a discrete set, H_A^S and H_B^S are discrete random variables. According to (9), we obtain the following:

$$H_A^S(\omega_1) = 1,\ H_A^S(\omega_2) = 0.9,\ H_A^S(\omega_3) = 0.6,\ H_A^S(\omega_4) = 0.5,\ H_A^S(\omega_5) = 0.3, \text{ and } H_A^S(\omega_6) = 0.1,$$
$$H_B^S(\omega_1) = 1,\ H_B^S(\omega_2) = 0.84,\ H_B^S(\omega_3) = 0.36,\ H_B^S(\omega_4) = 0.2,\ H_B^S(\omega_5) = 0.5, \text{ and } H_B^S(\omega_6) = 0.8.$$

Both the random variables H_A^S and H_B^S taking on these values with the fuzzy probabilities expressed by triangular fuzzy numbers, whose significant values equal to $\langle 1/12, 1/6, 3/12\rangle$.

According to (10) and (11), the expected values EH_A^S and EH_B^S coincide with EH_A^Z and EH_B^Z, which are expressed by triangular fuzzy numbers with the significant values given by Table 3. The significant values of the variances $var\ H_A^S$ and $var\ H_B^S$, computed by (12) and (13), are shown in Table 3.

Characteristic	Fuzzy Yield (in %)			Centre of Gravity	Middles of Maximum
$EH_A^Z = EH_A^S$	0.43	0.57	0.70	0.567	0.57
$EH_B^Z = EH_B^S$	0.49	0.62	0.75	0.617	0.62
$var\ H_A^Z$	0.08	0.12	0.15	0.118	0.12
$var\ H_B^Z$	0.10	0.13	0.15	0.127	0.13
$var\ H_A^S$	0.06	0.09	0.13	0.097	0.09
$var\ H_B^S$	0.05	0.08	0.10	0.078	0.08

Table 3 Expected Values and Variances of the Considered Random Variables

The resultant fuzzy expected values and fuzzy variances can be compared on the basis of various characteristics that are proposed in the literature, like centres of gravity or middles of maximums. The selection of the method of comparison is case sensitive. For illustration, let us employ the two mentioned characteristics. The *centre of gravity* of a fuzzy number A, whose support belongs to $\mathbb{R}$, is a real number $cog(A)$ given as follows:

$$cog(A) = \frac{\int_{-\infty}^{\infty} \mu_A(x) \cdot x \, dx}{\int_{-\infty}^{\infty} \mu_A(x) \, dx}.$$

The middle of maximum of a fuzzy number A represents the centre of its core. It is the $Core\ A$ itself for the triangular number A. By applying the rule of maximization of the expected value and minimization of the variance, according to the centres of gravity of computed fuzzy characteristics of random variables H_A^Z and H_B^Z we would not be able to decide between the projects A and B because $cog\left(EH_A^Z\right) < cog\left(EH_B^Z\right)$ and $cog\left(var\ H_A^Z\right) < cog\left(var\ H_B^Z\right)$. We can see that in this case, since $cog\left(EH_A^S\right) < cog\left(EH_B^S\right)$ and $cog\left(var\ H_A^S\right) > cog\left(var\ H_B^S\right)$, the project B would be selected. The ordering of stocks based on their middles on maximums leads to the same results. However, as it was stated in Section 2, the interpretation of the characteristics EH_i^Z and $var\ H_i^Z$, $i = A, B$, is questionable. Thus, the possible selection based on these characteristics would not be trustworthy.

4 Conclusion

We have dealt with the problem of extension the decision matrix to the case of fuzzy states of the world and underlying fuzzy probability measure. We have extended the approach to the problem proposed by Talašová and Pavlačka [8] that is based on applying the Zadeh's probabilities of the fuzzy states of the world to the case of fuzzy probabilities of the fuzzy states of the world. We have found out that the meaning of obtained characteristics of the outputs of alternatives, namely the expected values and variances, is questionable. Therefore, we have introduced an another approach, consisting in deriving fuzzy rule-based systems from a decision matrix with fuzzy states of the world and underlying fuzzy probability measure. In such a case, the obtained characteristics of outputs, based on which the alternatives are compared, are clearly interpretable. We have proved that the resulting expected values of consequences are for both the approaches the same. In numerical example we have shown, that the variances differ, and that the final ordering of the alternatives according to both the approaches can be different.

Next research in this field could be focused on the cases, where the outputs of alternatives are given by fuzzy numbers, and/or the continuous fuzzy probability measure.

Acknowledgements

Supported by the grant IGA_PrF_2017_019 *Mathematical Models* of the *Internal Grant Agency of Palacký University Olomouc*.

References

[1] Chankong, V., Haimes, Y.Y.: *Multiobjective decision making: theory and methodology.* Elsevier Science B.V., New York, 1983.

[2] Torra, V., Narukawa, Y.: *Modeling Decisions: Information Fusion and Aggregation Operators*. Springer, Berlin Heidelberg, 2007.

[3] Pavlačka, O., Rotterová, P.: Probability of fuzzy events. In: *Proceedings of the 32nd International Conference Mathematical Methods in Economics* (J. Talašová, J. Stoklasa, and T. Talášek, eds.), Palacký University Olomouc, Olomouc, 2014, 760–765.

[4] Pavlačka, O., Rotterová, P.: Fuzzy Decision Matrices Viewed as Fuzzy Rule-Based Systems. In: *MME2016: Conference Proceedings* (A. Kocourek, and M. Vavroušek, eds.), Technical University of Liberec, Liberec, 2016, 641–644.

[5] Rotterová, P., Pavlačka, O.: New Approach to Fuzzy Decision Matrices. Submitted to *Acta Polytechnica Hungarica*.

[6] Rotterová, P., Pavlačka, O.: Probability of Fuzzy Events and Their Use in Decision Matrices. *Int. J. Mathematics in Operational Research* **9**, 4 (2016), 423–435.

[7] Sugeno, M.: *Industrial applications of fuzzy control.* Elsevier Science Pub. Co., New York, 1985.

[8] Talašová, J., Pavlačka, O.: Fuzzy Probability Spaces and Their Applications in Decision Making. *Austrian Journal of Statistics* **35**, 2&3 (2006), 347–356.

[9] Yoon, K.P., Hwang, Ch.: *Multiple Attribute Decision Making: An Introduction*. SAGE Publications, California, 1995.

[10] Zadeh, L.A.: Probability Measures of Fuzzy Events. *Journal of Mathematical Analysis and Applications* **23**, 2 (1968), 421–427.

On the relation of labour productivity, costs and unemployment in the EU countries

Lenka Roubalová[1], Tomáš Vaněk [2], David Hampel[3]

Abstract. This paper deals with the analysis of the relationship between unit labour costs, real labour productivity and unemployment rate in the EU countries. The main objective is to determine the existence and the direction of Granger causality between labour productivity and unit labour costs. In addition, we focus on the relationship between these two variables and unemployment. To find out the causality direction, the VAR models are estimated. Impulse response analysis is used to describe the time path of one of the variables in the VAR to shocks from the other explanatory variables. Based on responses groups of similar countries are created. Our results prove the Granger causality running from labour productivity to unit labour costs just in Sweden, Finland, Luxembourg and Austria. The causality running from unit labour costs to labour productivity is observed in most of the rest of the European countries. In the case of Romania, Hungary and Cyprus, the relationship in the sense of Granger causality is not significant.

Keywords: EU, Granger causality, labour costs, labour productivity, unemployment, VAR model.

JEL classification: 91B84
AMS classification: C32

1 Introduction

Productivity growth is one of the most important factors for long-term economic growth and country competitiveness sustaining. There are many areas in the economics closely related to the productivity. In [23] we learn productivity etc. are important variables in assessment of regional disparities. Productivity significantly affects real exchange rates in the Eurozone [17], as well as in the Czech Republic, see [1]. As the theoretical literature and many empirical research articles show, changes in productivity development are associated with the movements in real wages, unemployment rate or inflation. At the macroeconomic level, it is very important to find out the interrelationship among these variables, especially for policy makers and other authorities responsible for several reforms and political decisions in this field. Moreover, at the microeconomic level the results of such analysis can be beneficial for management decisions. OECD presents in [16] that the productivity issue connected with living standards increasing and economic growth in general has become very popular among policy makers as a consequence of the crisis in 2008.

The economic theory in this case is not uniform. The neoclassical approach based on the statement that wages equal productivity as measured by the value of the marginal product of labour results from the labour demand analysis (connected with marginal productivity theory). The wages depend on workers' productivity, so highly productive workers are highly payed, see [13]. In accordance with this theory Mankiw in [13] describes the labour productivity development in USA during 1973–1995. The labour productivity growth had been followed by the increase in real wages. Conversely, the efficiency wage theory states that causality runs from real wages to productivity, which is connected with the thoughts of Yellen, see [21]. The theory rejects the assumption that the previous theory is based on. According to the efficiency wage theory, the wage setting at the level above the market wage level has positive effect on workers' efforts increasing their productivity and efficiency. This is the way how firms attract experienced workers. Wage rigidity and involuntary unemployment can be explained by this theory as well, because job seekers are willing to wait for well-payed job and such behaviour leads to this phenomenon.

Many authors focus on analysing the relationships among labour productivity, real wages (or labour costs) and unemployment (or alternatively inflation). The link between this issue and inflation is described in [5] or [18]. Conversely, the influence of EU labour market institutions on total factor productivity is in [4]. As Goh in [9] argues, wage development is affected by both, productivity and unemployment. We can find an empirical evidence in behalf of the efficiency wage theory in [20], the author proves negative impact of wages to labour productivity, but reverse causality is not proved. Yildirim, see [22], focuses on the relationship among inflation rate, labour productivity and wages in the processing industry in Turkey. Kumar, see [12], analyzes the same relationship

[1]Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, xkravack@node.mendelu.cz
[2]Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, xvanek7@node.mendelu.cz
[3]Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, david.hampel.uso@mendelu.cz

as Yildirim does and he proves the impact of wages to the labour productivity in the long run in Australia. The causality running from labour productivity to wages is proved in [19] and in the work of Klein, see [11], who proves long-term impact of labour productiivty to wages in South Africa.

The main objective of this paper is to determine the existence and direction of Granger causality among labour productivity and unit labour cost in the EU countries in the period 2001–2016. In addition, we focus on the interrelationship of these two variables and unemployment in these countries. Finally, based on responses similar countries are grouped.

2 Material

In the empirical analysis we use variables as follows: unemployment rate defined as a share of unemployed persons and economically active population (UNP), real labour productivity based on hours worked (LP) and real unit labour costs based on hours worked (ULC) representing the cost of labour as a producing factor. The empirical analysis also takes into account the possibility of alternative using real labour productivity and unit labour costs based on persons employed, but the results obtained by this way are almost in all cases in line with the previous ones. However, this data expression is used for Belgium and Croatia, because the dataset based on hours worked is not complete for these countries. We use set of quarterly time series data, where the number of time series observations $T = 63$ corresponds to the period Q1 2001 – Q3 2016 for each country. The data are expressed as an index, change against the same quarter of the previous year because of the availability of the data at the desired length. In addition, this expression is convenient from the point of view of seasonally adjustment. Then, all time series are stationarized by using first differences. For time series with very volatile index of variables during observed period, using growth factors instead of first difference seems to be more appropriate. We revealed that both approaches lead to the same results in our case, so we present the estimated models based on using time series of the first differences. All datasets are obtained from the Eurostat database, see [6], [7].

3 Methods

To find out the causality direction among all variables the vector autoregressive (VAR) models are estimated. VAR is special case of multivariate time series model. Such time series are created as a combination of at least two time series. This representation is advantageous when there exists dependency between components of one time series and between time series of each other. All VAR models estimated in this article are created as a combination of three time series. VAR model can be formulated as

$$\boldsymbol{X}_t - \Phi_1 \boldsymbol{X}_{t-1} - \Phi_2 \boldsymbol{X}_{t-2} - \cdots - \Phi_p \boldsymbol{X}_{t-p} = \boldsymbol{Z}_t, \tag{1}$$

where $\Phi_1, \ldots, \Phi_p$ are matrices of coefficients of the size 3×3, $\boldsymbol{X}_t, \ldots, \boldsymbol{X}_{t-p}$ are vectors of variables up to lag p of the size 3×1 and $\boldsymbol{Z}_t$ is vector of random errors of the same size.

Optimal lag length for the VAR model we establish using autocorrelation and partial autocorrelation function and information criteria (AIC, BIC, HQC). The decision about the most appropriate lag length is supported by using Q test [3] for VAR model and t-test for the estimated parameters of relevant lag. Then the residuals are tested to verify the white noise character. Autocorrelation is detected with using Ljung-Box test. If there is evidence of autocorrelation, more lags should be added to eliminate it. Doornik-Hansen, Lilliefors and Jarque-Bera tests are used to judge normality of residuals.

To detect the Granger causality, F-tests are used. The VAR model can be divided into three equations where each of them is relevant for one variable. So each variable is estimated as a combination of lagged parameters of that variable and of the two others. F-test is used to prove the significant influence of the lagged values of selected parameters to the variable which the equation is estimated for. The null hypothesis states, that all parameters of concrete variable are equal to zero (not significant). When H_0 is rejected, the variable can be seen as causal in the sense of Granger causality to the variable the equation is estimated for. Then, impulse response analysis is used as a tool to describe the response time path of one of the variables in the VAR to shock from the other variable. Based on these responses, groups of similar countries are created. For all tests we use significance level $\alpha = 0.05$. All calculations are performed in software Gretl, version 2016d and Matlab, version R2016b.

4 Results

Obtained results provided by the analysis are summarized in Table 1. The causality running from labour productivity to labour costs is detected in the case of Sweden (SE), Finland (FI), Luxembourg (LU) and Austria (AT). The impulse response analysis shows that an increase in the labour productivity leads to the rise of unit labour costs and then it shows profound fluctuations in the long run (Figure 1, E). In the case of Luxembourg and Finland the unit labour costs response is the same but rather temporary.

Country	VAR(p)	LP → ULC	ULC → LP	ULC → UN	UN → ULC	LP → UN	UN → LP
AT	8	0.0101	< 0.0001	×	×	×	×
BE	5	×	0.0003	×	×	0.0410	×
NL	6	×	0.0016	0.0057	×	×	×
DK	6	×	0.0355	×	×	0.0012	×
DE	1	×	< 0.0001	0.0101	×	×	×
LU	4	0.0221	×	×	×	×	×
IE	5	×	0.0128	×	×	×	×
UK	4	×	×	×	0.0120	×	×
FR	4	×	0.0066	×	×	×	×
PT	4	×	0.0020	×	×	×	×
ES	5	×	0.0258	×	×	×	×
IT	6	×	×	0.0256	×	×	×
GR	5	×	×	0.0304	×	×	×
CY	5	×	×	×	×	×	×
MT	9	×	0.0153	×	×	×	×
HR	3	×	×	×	×	×	0.0037
SI	7	×	0.0044	×	×	×	×
BG	3	×	×	< 0.0001	×	×	×
RO	5	×	×	×	×	×	×
CZ	7	×	0.0165	×	×	×	×
SK	12	×	×	0.0156	×	×	×
PL	8	×	×	0.0247	×	0.0070	×
HU	9	×	×	×	×	×	×
LV	7	×	0.0074	×	×	×	×
LT	7	×	0.0130	×	×	×	×
EE	9	×	×	0.0015	×	×	×
FI	6	0.0359	×	×	×	×	×
SE	8	0.0188	×	×	×	×	×

Table 1 All EU countries are marked by ISO code. The columns show relationship in the sense of Granger causality and its direction among LP, ULC and unemployment rate UNP. For significant dependency among these variables, the p-values of F-test are shown. If there is $\times$ instead of p-value, there is no significant causality.

The results for most of the EU countries are in favour of the efficiency wage theory. Granger causality running from unit labour costs to labour productivity is detected in the case of Austria, Belgium (BE), the Netherlands (NL), Denmark (DK), Germany (DE), Ireland (IE), France (FR), Portugal (PT), Spain (ES), Malta (MT), Slovenia (SI), Czech Republic (CZ), Latvia (LV), and Lithuania (LT).

Impulse response analysis presents that an increase in the unit labour costs causes sharp decline in labour productivity followed by short-time increase. So finally, after the short time oscillation the LP returns to the original level in the case of Denmark, the Netherlands, Germany, Spain, Portugal, Malta and the Czech Republic (Figure 1, A). Labour productivity response for Slovenia, Austria, Latvia, Lithuania, Belgium and Ireland is similar, but only the reaction is not only temporary (Figure 1, B). In the case of France the shock in unit labour costs leads to the same change of labour productivity, but finally it rather declines. Table 1 also shows there is no causality between unit labour costs and labour productivity (and vice versa) in the Great Britain (UK), Greece (GR), Cyprus (CY), Croatia (HR), Italy (IT) Bulgaria (SL), Romania (RO), Slovakia (SK), Poland (PL) Hungary (HU) and Estonia (EE).

Apart from labour productivity and unit labour costs relationship, the analysis proves relationship of these variables and unemployment. For some of the countries with no causality between labour productivity and unit labour costs (and vice versa), the Granger causality running from unit labour costs to unemployment rate is proved as well as it is proved for some countries with significant Granger causality between labour productivity and unit labour costs. This is the case of the group of countries including the Netherlands, Germany, Italy, Greece, Bulgaria, Slovakia, Poland and Estonia.

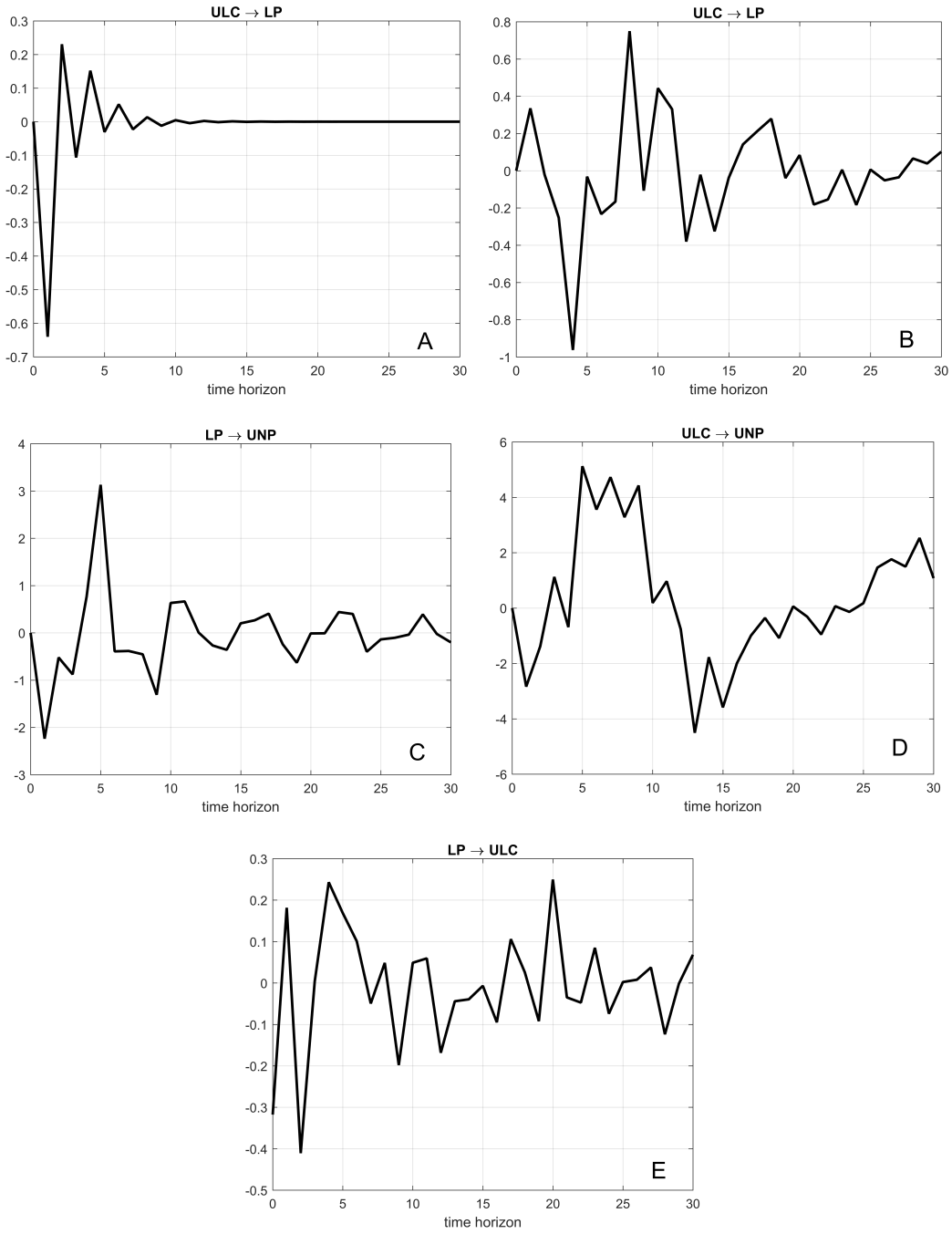


Figure 1 Graphs A–E describing reaction of one variable to the shock change in another one. Each graph represents relationship typical for each group of countries represented by one of them. Concretely, A – Germany, B – Latvia, C – Denmark, D – Estonia, E – Sweden.

For Denmark, Poland and Belgium, the causality running from labour productivity to unemployment rate is found. As impulse response analysis shows, shock increase in labour productivity is followed by unemployment decline first, then it rises and finally continues with slight oscillations (Figure 1, C).

Impulse response analysis also shows, unemployment rate declines in some short time as a result of unit labour cost increase. Then, unemployment rate increases sharply, oscillates at the higher level, then declines and slowly increases in the case of Estonia (Figure 1, D), in the case of Poland the decrease is very slight and in the case of the rest of the countries the unemployment rate returns back and stabilizes at the equilibrium level.

There are also some peculiarities resulting from the analysis. There is causality running from unemployment rate to labour productivity in Croatia (HR) and from unemployment rate to unit labour costs in the Great Britain

(UK). As the unemployment rate increases, labour productivity as well as unit labour costs decline slowly. There are also countries where no causality is detected among analyzed variables. These countries are Hungary (HU), Romania (RO) and Cyprus (CY).[4]

5 Discussion

Findings resulting from our analysis state, that in the most of EU countries there is causality running from labour costs to labour productivity. These results are in accordance with efficiency wage theory and with evidence in favour of this theory for South Africa [20], Turkey [22], Australia [12] and Canada, USA and Italy [8], but there is lack of similar analysis for EU countries. Authors focus on these countries in [15] or [19] but from the point of view of correlation and dynamics of development rather than Granger causality.

Granger causality running from labour productivity to labour costs is proved for the group of rich, highly advanced and economically stable countries including Sweden, Finland, Austria and Luxembourg[5]. The major group of countries can be characterized by Granger causality running from unit labour costs to labour productivity. According to Rusinova [19] the group of countries with this direction of causality mainly consists of countries with highly regulated wage bargaining system (Denmark, France, Germany, the Netherlands, Portugal) and with regulated system of bargaining (Belgium, Spain, Slovenia). As the author states, this group can be characterized by the existence of high level of collective agreement coverage, a dominance of sectoral wage bargaining.

Conversely, comparing with the same analysis of Rusinova [19], some of the countries with no causality between labour productivity and unit labour costs can be characterised by deregulated system of wage bargaining (UK, Estonia, Poland). The limitation is that her analysis focuses on 19 OECD countries, so it is not possible to find the context more deeply.

There is empirical evidence that may help to explain results revealing no causal relationship among all three variables. In [10] the relationship among these variables in Romania in 1991–2006 is investigated. According to the authors the high rate of labour market inflexibility is caused by wage inflexibility resulting from no correlation of real wage and labour productivity dynamics. The labour market in Romania is more rigid than any other markets. Apart from the economic factors, strong social factors interfere on this market so they consider the development of these variables to be atypical in Romania. Generally, similar results (proving no causality) for Romania, Cyprus and Hungary can be caused by similar impacts of the harmful effects of the financial or economical crisis. Different results for Croatia can be connected with the fact presented in [2] that regional impact on the Croatian labour market is considerable as compared with other EU countries.

There are not many relevant publications focused on this issue for EU countries so the possibilities of comparing and discussing them are limited. On the other hand, the lack of similar studies creates new possibilities for further research. It could be concentrated on the same interrelationships by using another methods of Granger causality detecting or it could concentrate on the economic influences affecting the direction of Granger causality and explaining why the results are just that. Generally, more authors should explore these relationships at the macroeconomic level.

6 Conclusion

The main aim of this article was to investigate the relationship between labour productivity and unit labour cost. In addition, we focused on the relationship between these two variables and unemployment in the EU countries in 2001–2016. To reach these results VAR models were estimated. Several statistical tests were performed to make sure the most suitable VAR model for each country was estimated. The results revealed the causality running from unit labour costs to labour productivity for the major group of EU countries and they were in accordance with the efficiency wage theory. Results obtained in the empirical part of this article were discussed and the discussion showed there was lack of similar articles concentrated on the causal interrelationship between these three variables in the EU 28.

Acknowledgements

This research is supported by the doctoral grant No. $PEF-DP-2017028$ of the Grant Agency IGA PEF MENDELU.

[4] Alternatively, the dataset using LP and ULC based on person employed is used also in the case of Ireland, Cyprus, Romania and Hungary to verify the previous results proving no causality. With the exception of Ireland the results obtained here are identical.

[5] Except of Luxembourg (historical member) the three countries joined the EU at the same time.

References

[1] Adámek, E.: The Role of Productivity in the Czech Republic: Does Balassa- Samuelson Effect Still Exist? In: *Proceedings of the 34th International Conference Mathematical Methods in Economics*, (2016), 1–6.

[2] Botric, V.: Regional Aspects of Unemployment in Croatia. *Privredna kretanja i ekonomska politika (Economic Trends and Economic Policy)* **95** (2003), 27–46.

[3] Box, G., and Pierce, D.: Distribution of Residual Autocorrelations in Autoregressive Moving Average Time Series Models. *Journal of the American Statistical Association* **65** (1970), 1509–1526.

[4] Čekmeová, P.: Labour Market Institutions and Total Factor Productivity - An Evidence in the European Union. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics*, (2016), 108–113.

[5] Dobešová, A., and Hampel, D.: Asymmetries in Effects of Monetary Policy and their Link to Labour Market - Application of TVP-VARs to Three European Countries. In: *Proceedings of the 32nd International Conference Mathematical Methods in Economics*, (2014), 156–161.

[6] Eurostat: *Labour productivity and unit labour costs* [online], (2017). Available at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nama-10-lp-ulc&lang=en.

[7] Eurostat: *Unemployment rate, quarterly data, seasonally adjusted* [online], (2017). Available at: http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tipsun30&plugin=1.

[8] Fuess, S., and Millea, M.: Pay and Productivity in Corporatist Germany. *Journal of Labor Research* **27** (2006), 397–409.

[9] Goh, S. K.: Is Productivity Linked To Wages? An Empirical Investigation in Malaysia. *CenPRIS Working Paper Series* **9** (2009), 1–19.

[10] Herman, E., and Georgescu, M. A.: Correlations Between the Average Wage and Labor Productivity in Romania in the context of the socio-economic sustainable development. In: *Proceedings of the International Conference on Economics, Law and Management*, (2009), 34–43.

[11] Klein, N.: Real Wage, Labor Productivity, and Employment Trends in South Africa: A Closer Look. *IMF Working Paper*, WP/12/92: (2012), 1–28.

[12] Kumar, S.: Real wages, Inflation and Labour Productivity in Australia. *Applied Economics* **44** (2012), 2945–2954.

[13] Mankiw, G., N. *The Principles of Economics*. South-Western Cengage Learning, Mason, 2012.

[14] Meager, N., and Speckesser, S.: Wages, Productivity and Employment: A review of theory and international data. *European Employment Observatory* (2011), 51–73.

[15] Ninkulin, D.: Relationship Between Wages, Labour Productivity and Unemployment Rate in New EU Member Countries. *Journal of International Studies* **8** (2015), 31–40.

[16] OECD: Compendium of Productivity Indicators 2013, OECD Publishing [online]. (2016). Available at: http://dx.doi.org/10.1787/pdtvy-2013-en.

[17] Plecitá, K., and Issever Grochová, L.: Productivity and Real Exchange Rates in the Euro Area. In: *Proceedings of the 33rd International Conference Mathematical Methods in Economics*, (2015), 659–664.

[18] Roubalová, L., and Hampel, D.: The Main Inflationary Factors in the Visegrád Four. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics*, (2016), 741–746.

[19] Rusinova, D., Lipatov, V., and Heinz, F. F.: How Flexible Are Real Wages in EU Countries? A panel investigation. *Journal of Macroeconomics* **43** (2015), 140–154.

[20] Wakeford, J.: The Productivity-wage Relationship in South Africa: an empirical investigation. *Development Southern Africa* **21** (2004), 109–132.

[21] Yellen, J.: Efficiency Wage Models of Unemployment. *The American Economic Review* **74** (1984), 200–205.

[22] Yildirim, Z.: Relationships among labour productivity, real wages and inflation in Turkey. *Economic Research-Ekonomska Istraživanja* **28** (2015), 85–103.

[23] Žižka, M.: Model of Evaluation of Socio-Economic Disparities of Sub-Regional Units in the Czech Republic. In: *Proceedings of the 32nd International Conference Mathematical Methods in Economics*, (2014), 1130–1135.

Entropy Based Measures Used in Operation Complexity Analysis of Supplier-Customer Systems

Pavla Říhová[1], Ladislav Lukáš[2]

Abstract. The paper deals with operational complexity of supplier-customer systems. Measurement of operational complexity of company supplier-customer relations belongs to important managerial tasks. In the present work, we deal with quantitative measuring of these relations on the base of unified entropy-oriented approach. From analytic point of view, it reflects temporal changes in supplier-customer system. Operational complexity measure should express behavioral uncertainties of the system during the time with respect to specified control levels. It has to record all possible types of flow variations within and across company in detail, e.g. replenishment time disturbances, deviations of material in/out flows, etc. Further, we discuss in detail various entropy measures which are applicable to operational complexity analysis. The basic one is classical Shannon entropy. We also discuss relations between the Shannon entropy and various generalized entropy based measures, e.g. (c,d)-entropy. We present numerical results of investigation of equivalent classes between various entropy measures. All computations are performed in sw Mathematica, and algorithmic details are presented.

Keywords: entropy, generalized entropy, information measure, operational complexity, supplier-consumer system.

JEL classification: C63, C81
AMS classification: 90B99

1 Introduction

In general, supplier-customer systems can be handled from various viewpoints and using different approaches, and interactions among suppliers and customers can be rather complex. The extant literature provides many important but frequently too specific views of the system supplier's roles and customer relationship dynamics. Business economics knows basically two types of supplier-customer system complexity – structural one and operational one. The first one represents more-less a static viewpoint concerning with structural links and relations among system units or entities mainly, whereas the operational complexity is focused on all uncertainties and perturbations within dynamic evolution of the system.

From an analytic point of view, operational complexity reflects temporal changes in supplier-customer systems. In particular, an operational complexity measure should express the behavioral uncertainties of the system during time with respect to specified control levels. It has to record all possible types of flow variations within and across companies in detail, e.g., replenishment time disturbances, deviations of material in/out flows, etc. We assume that such data are available in a company management information system (MIS).

In our previous papers [3] and [4], we presented both details of information schemes of supplier-customer systems and basic structure of the problem-oriented database on which the operational complexity algorithms based upon entropy are operating. In particular in [4], we also discussed some relations between well-known Shannon entropy, or the Boltzmann-Gibbs-Shannon one as it is alternatively called (denoted as BGS-entropy, too), and a more general entropy, the (c, d)-entropy, in particular. In this paper, we concern with several other entropies and entropic measures which can be used for construction of supplier-customer operational complexity measures emphasizing various managerial aspects prospectively.

In [1] and [10], in particular, there are discussed general entropy-based approaches for measuring complexity in supply chains and supplier-customer systems, too. Within tremendous literature devoted to generalized entropy and entropic measures available, we select such ones seeming us most contributable to the paper topic. In [5], the interesting novel entropic measure is given which combines nonextensivity of Tsallis entropy and additivity property of Renyi entropy. In [9], based upon Tsallis entropy, the new generalized sample entropy is given and applied to traffic signals analysis. The multidimensional generalized entropy measure is given in [6] and [8], and it

[1]University of West Bohemia, Faculty of Economics, Dept. of Economics and Quantitative Methods, Univerzitní 22, 306 14 Pilsen, divisova@kem.zcu.cz
[2]University of West Bohemia, Faculty of Economics, Dept. of Economics and Quantitative Methods, Univerzitní 22, 306 14 Pilsen, lukasl@kem.zcu.cz

provide us a good platform for research and construction of operational complexity measuring algorithms in case when supplier-customer business contracts contain replenisments with undecomposable bundles of products or services. Finally, in [2], the generalized (c, d) entropy is thoroughly discussed, together with its links to other known entropies, the Shannon entropy, in particular. Some aspects of the (c, d) entropy are given in [4], too, whereas the α-power entropy has been already discussed in [3], with some numerical experiments and the interesting quadratic entropy, as a particular case. Finally, in [7], pure mathematical derivation of classic entropy, i.e. BGS-entropy, is given as a solution of specific equation.

Let $P = \{p_1, p_2, \ldots, p_n\}$, $0 \leq p_i \leq 1, i = 1, \ldots, n$, $\sum_{i=1}^{n} p_i = 1$, is a discrete probability distribution of n states of a system, then the BGS-entropy of P is

$$S_{\mathrm{BGS}}(P) = -\sum_{i=1}^{n} p_i \log(p_i), \tag{1}$$

where the original Shannon entropy, being developed within the information theory, was proposed to use logarithmic function with base $2, \log_2$, whereas the usual form of BGS-entropy prefers natural logaritms.

All our previous case studies handled operational complexity analyses of various supplier-customer systems of SME ranked firms and published in [3] and [4], were based upon the BGS-entropy (1), in principle. However, before presenting the other entropies and entropic measures, we overview a basic information framework of supplier-customer systems supporting operational complexity analyses.

We consider a set of products $\{\Pi_1, \ldots, \Pi_m\}$ within a supplier-customer system being specified by a business contract between both sides for a given time period. In general, for any product $\Pi_j, j = 1, \ldots, m$, we may consider two types of variables relating quantity and time, which provide corresponding time series monitored and stored in company MIS, and denoted $\{_{(\mathrm{e,r})}Q_j\}$, and $\{_{(\mathrm{e,r})}T_j\}$, respectively. Let any $_{(\mathrm{e,r})}Q_j$, and $_{(\mathrm{e,r})}T_j$, contain particular entries denoted $_{(\mathrm{e,r})}q_{j,k}$ and $_{(\mathrm{e,r})}t_{j,k}$, for $k = 1, \ldots, \kappa_j$, with κ_j given. Couple of pre-indeces (e,r) is to specify logical relations in supplier-customer system, in particular $\mathrm{e} \in \{\mathrm{s}, \mathrm{i}, \mathrm{c}\}$, s... supplier, i... interface, c... customer, and $\mathrm{r} \in \{\mathrm{f}, \mathrm{o}, \mathrm{d}\}$, f... forecast, o... order, d... delivery,

Let $\{\mathcal{A}_1, \ldots, \mathcal{A}_n\}$ denote discrete system states selected for a product Π_j and its quantity or time variable particularly. In general, noting it precisely to be a rather crucial problem because first, being product-depedent as usual, and second, it should reflect all prospective aspects and various managerial purposes of operational complexity analyses. Using the discrete system states $\{\mathcal{A}_1, \ldots, \mathcal{A}_n\}$ and corresponding data available, we calculate the very first issue of the data processing phase, i.e. empirical distribution $P = \{p_1, \ldots, p_n\}$. Further, this one serve us for entropy based operational complexity analysis. Operational complexity measure should express behavioral uncertainties of the system during the time with respect to specified control levels. It has to record all possible types of flow variations within and across company in detail, e.g. replenishment time disturbances, deviations of material in/out flows, etc.

2 Some entropies and entropic measures

First, we discuss some entropies given in [2] and [5]. Renyi and Tsallis entropy are both generalization of BGS-entropy, i.e. Shannon entropy. Renyi entropy is additive for statistically independent subsystems for entropic parameter α, and it can handle the additive information that exist in system. Tsallis (nonextensive) entropy has pseudo-additivity property with entropic parameter β, sometimes denoted q and called Tsallis index, too. It can handle a nonextensive information for statistically independent subsystems through its pseudo-additivity property.

The *Renyi entropy* of order α is defined for any discrete probability distribution $P = \{p_1, p_2, \ldots, p_n\}$, $0 \leq p_i \leq 1, i = 1, \ldots, n$, $\sum_{i=1} np_i = 1$, as follows

$$H_\alpha(P) = \frac{1}{1-\alpha} \log\big(\sum_{i=1}^{n} p_i^\alpha\big), \quad \alpha \in (0,1) \cup (1, +\infty). \tag{2}$$

In case $\alpha = 1$, the Renyi α-entropy $H_1(P)$ is the classical BGS-entropy.

A link between Renyi α-entropy and Tsallis β-entropy is established through the bijective function

$$\phi_\alpha(z) = \exp(z - \alpha z),\ \phi_\alpha^{-1}(y) = \frac{1}{1-\alpha} \log(y), \quad \alpha \in (0,1) \cup (1, +\infty),\ \forall z, y \in \mathcal{R}_+. \tag{3}$$

The *Tsallis entropy* of order β is defined for any discrete probability distribution $P = \{p_1, p_2, \ldots, p_n\}$, $0 \leq p_i \leq 1, i = 1, \ldots, n$, $\sum_{i=1} np_i = 1$, as follows

$$S_\beta(P) = \frac{\phi_\beta(H_\beta(P)) - 1}{1-\beta} = \frac{1}{1-\beta}\big(\sum_{i=1}^{n} p_i^\beta - 1\big), \quad \beta \in (0,1) \cup (1, +\infty). \tag{4}$$

In the limit $\beta \to 1$, $S_\beta(P)$ recovers the classical BGS-entropy. Let us accept a temporal change of q instead of β in order to write (4) in another equivalent form using so called q-logarithmic function $\log_q$ in following way

$$S_q = \sum_{i=1}^{n} p_i \log_q(1/p_i)\big), \quad \log_q(x) := \frac{x^{1-q}-1}{1-q}, \quad x \in \mathcal{R}_+,\, q \in \mathcal{R}, \quad \log_1(x) := \ln(x). \tag{5}$$

Following [5], the novel generalized entropy, called *Masi entropy*, is defined. It combines the nonextensivity of Tsallis entropy and the additivity of Renyi entropy. The main difference of Masi enropy to Renyi and Tsallis entropies is that the entire probability function is raised to a power r, in contrast to the probability functions in Renyi and Tsallis entropy, where every state-probability separately is raised to a power α, or β, respectively. The Masi entropy is

$$S_r(P) = \frac{1}{1-r} \log\big(1-(1-r)\sum_{i=1}^{n} p_i \log p_i\big), \tag{6}$$

where $r > 0, r \neq 1$, and it is used to measure the degree of additivity/nonextesitivity in system.

From (6), we can see that for $r = 1$ the Masi entropy reduces to the BGS-entropy, i.e. $S_1(P) = -\sum_{i=1}^{n} p_i \log p_i$, in similar way as the Renyi entropy $S_{\alpha=1}(P)$, and the Tsallis entropy $S_{\beta=1}(P)$, does.

For two statistically independent systems described by the probability distributions $P = \{p_1, p_2, \ldots, p_n\}$, $Q = \{q_1, q_2, \ldots, q_n\}$, $i = 1, \ldots, n$, the Masi entropy satisfies

$$S_r(P \cap Q) = S_r(P) + S_r(Q), \tag{7}$$

which represents the additivity property of Masi entropy, and the parameter r can be viewed as a measure for the degree of nonextensivity that exists in system.

The generalized (c, d)-*entropy* is given in [2], and discussed also in [4]. It is defined by following expression

$$S_{c,d}(P) = er \sum_{i=1}^{n} \Gamma(d+1, 1-c\log p_i) - cr, \quad r = (1-c+cd)^{-1}, \tag{8}$$

here c, d are specific parameters, and $\Gamma(a, b)$ is the incomplete gamma function defined by relaxing lower integration bound of well-known Euler gamma function $\Gamma(a)$, as follows

$$\Gamma(a, b) = \int_b^{+\infty} t^{a-1} \exp(t) \mathrm{d}t, \qquad \Gamma(a) = \int_0^{+\infty} t^{a-1} \exp(t) \mathrm{d}t \tag{9}$$

noting that $\Gamma(a)$ provides a useful extension of the factorial function for any $a > 0$, as $(n-1)! = \Gamma(n)$ holds for any positive integer n.

Following [2], the relation between (c, d)-entropy and BGS-entropy, is established for $(c, d) = (1, 1)$, for any discrete probability distribution P.

$$S_{1,1}(P) = S_{\mathrm{BGS}}(P) - 1. \tag{10}$$

As already presented in [4], this relation had motivated us to propose a modified (c, d)-entropy with slightly different additive term

$$S^*_{c,d}(P) = er \sum_{i=1}^{n} \Gamma(d+1, 1-c\log p_i) - c(r+1), \quad r = (1-c+cd)^{-1}, \tag{11}$$

thus yielding relation (12), a little bit 'smarter' one than (10)

$$S^*_{1,1}(P) = S_{\mathrm{BGS}}(P). \tag{12}$$

In [6] and [8], there is presented a multi-dimensional generalized entropy measure, denoted GEM, but let us call it $\mathcal{G}$*-entropy*, which is suitable for measuring some sociological mobilities and variability in income levels. In supplier-consumer operational complexity analysis, the multi-dimensional generalized entropy measure can be used in any case when supplier deliveries contain undecomposable bundles of goods.

Consider a sample of N records containing quantity or time data of a bundle of K products, which we want to analyse jointly. The corresponding data available are collected in the matrix X, of dimension $N \times K$

$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots \\ \vdots & \ddots & \\ x_{N1} & \ldots & x_{NK} \end{bmatrix} \tag{13}$$

Following [6], the generalized $\mathcal{G}$-entropy measure of X is defined in following way

$$\mathcal{G}_\gamma(X) = \frac{1}{\gamma(1+\gamma)}\frac{1}{N}\sum_{i=1}^{N}\Big(\big(\frac{s_i}{\bar{s}}\big)^{1+\gamma}-1\Big), \quad \gamma \neq -1, 0,$$
$$s_i = \Big(\sum_{j=1}^{K}\delta_j x_{ij}^{-\beta}\Big)^{-1/\beta}, \quad i = 1,\dots,N, \qquad \bar{s} = \frac{1}{N}\sum_{i=1}^{N} s_i, \tag{14}$$

where γ represents an entropic parameter, s_i is generalized mean of order $-\beta$, $\bar{s}$ is arithmetic mean of s_i, and $\delta_j, j = 1,\dots,K$, $\delta_j \in [0,1]$, and β, $\beta \in [-1,+\infty[$, are additional parameters with specific meanings. In particular, δ_j is the weight of each variable/product j, and β is an internal data tempering parameter. In case $\gamma = -1, 0$, the $\mathcal{G}_{-1}(X), \mathcal{G}_0(X)$ are given as follows

$$\mathcal{G}_{-1}(X) = \frac{1}{N}\sum_{i=1}^{N}\big(\frac{\bar{s}}{s_i}\big), \quad \mathcal{G}_0(X) = \frac{1}{N}\sum_{i=1}^{N}\big(\frac{s_i}{\bar{s}}\big). \tag{15}$$

3 Numerical results

First, we investigate behavior of the Masi entropy $S_r(P)$, for given probability distribution P. Since the uniform distribution gives the maximal entropy, thus expressing the maximal uncertainty, naturaly, we select $P = U = \{p_i = 1/n\}, i = 1,\dots,n = 10$. The behavior of $S_r(U), r \in [.01, .99] \cup [1.01, 1.45)$, is depicted in Figure 1, and listed in Table 1, for a series of r approaching a limit value.

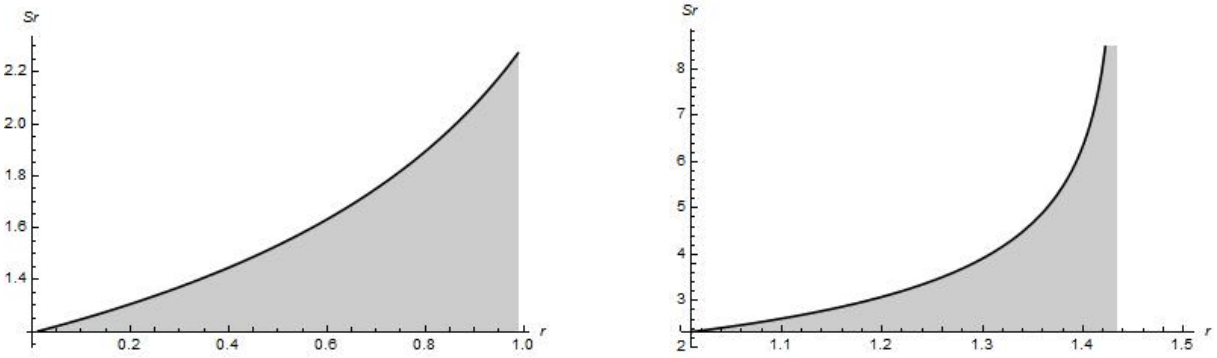


Figure 1 Masi entropy: *left panel*: $S_r(U), r \in (0,1)$; *right panel*: $S_r(U), r \in (1, 1.5)$

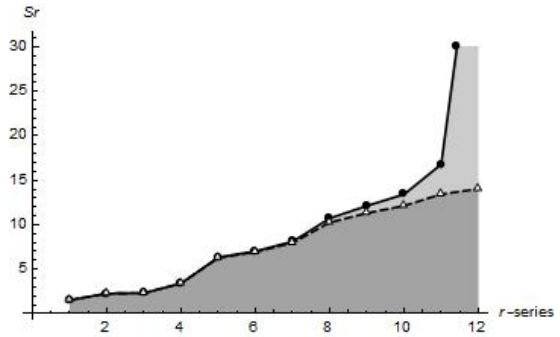


Figure 2 Masi entropy: $S_r(U)$ (full), $S_r(U_s)$ (dashed line), for r given in Table 1

case:	1	2	3	4	5	6	7	8	9	10	11	12
r	.5	.99	1.01	1.25	1.40	1.41	1.42	1.43	1.432	1.433	1.434	1.4342944815
S_r	1.53	2.28	2.33	3.43	6.35	7.03	8.13	10.76	12.14	13.43	16.81	47.8878

Table 1 S_r entropy of $U = \{p_i = 1/n\}, i = 1,\dots,n = 10$, for series of r approaching a limit

For given uniform distribution $U = \{p_i = 1/n\}, i = 1, \ldots, n = 10$, a steep increasing function $S_r(U)$, for $r > 1.4$, indicates an existence of a limit value. Using Mathematica, we computed $r = 1.4342944815$, as acceptable computational lower bound of the limit value $r_{\lim}$, such that $\lim S_r(U) = +\infty, r \to r_{\lim}$, as the value $S_{r=1.435}(U) = 14.7645 - 7.22205i$, is already complex.

Further, we also generated a sample uniform distribution U_{s} issuing 1000 integers by Mathematica function `RandomInteger` taking values $0, 1, \ldots, 9$, randomly. Calculating frequencies at individual bins $0, 1, \ldots, 9$, representing system states prospectively, we obtained the distribution U_{s}, given in Table 2. The comparison of $S_r(U)$ and $S_r(U_{\mathrm{s}})$ is depicted in Figure 2.

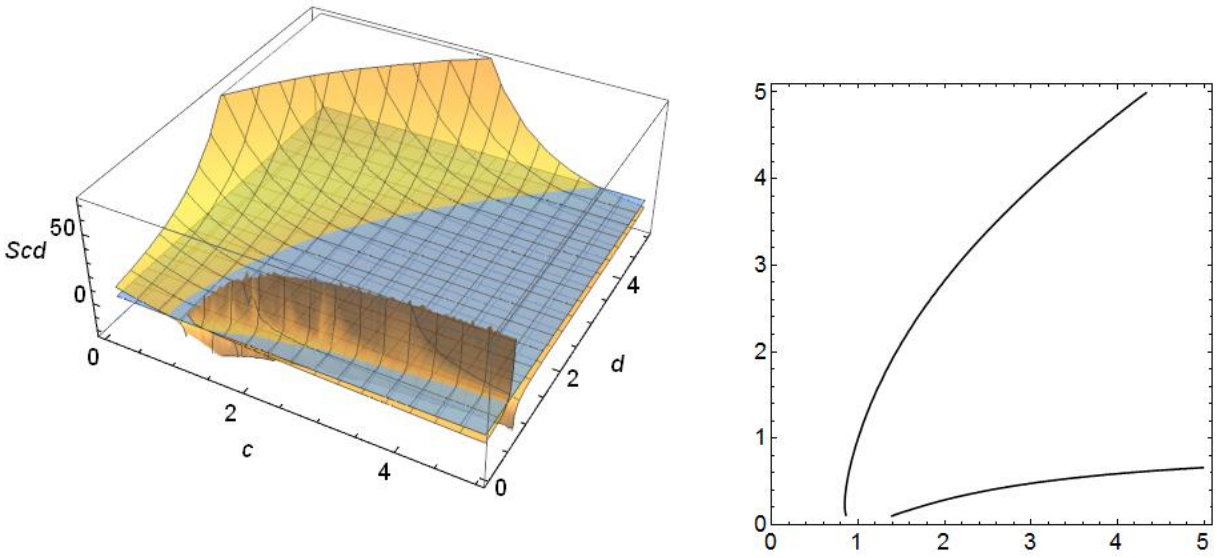


Figure 3 The iso-value curve $\lambda_1(c, d) = 0$ in 2-D defined by Eq.(16) – *left panel*: 3-D view

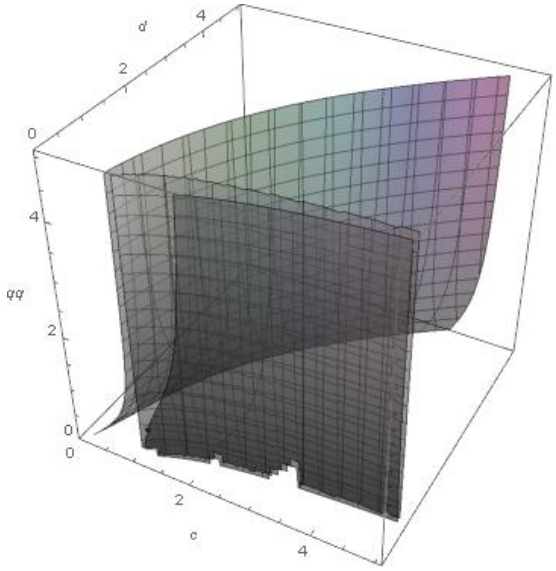


Figure 4 The 2-D iso-value manifold $\lambda_2(c, d, q) = 0$ in 3-D defined implicitly by Eq.(17)

The Masi and Tsallis entropies, S_r and S_β, but denoted q, here, are computed by Mathematica pure functions

```
Sr := Log[1 - (1 - #) Total[Table[piDistr[[i]] Log[piDistr[[i]]],
       {i, n}]]]/(1 - #) &
sQ := (1 - Total[Table[pidata = data[[i]] recipNlist; pidata^#,
        {i, imax + 1}]])/(# - 1) &
```

here the probability distribution $P = \{p_1, p_2, \ldots, p_n\}$ is set in array `piDistr`, or `pidata`, respectively, and formal argument denoted # is substituted by value of parameter r, or q, when invoking the functione, for example `Sr[.5]`, or `sQ[.5]`.

The generalized (c, d)-entropy is computed by following Mathematica module

```
scd[c_, d_, data_, recipN_, imax_, iswitch_] :=
 Module[{cc = c, dd = d, pi, rr, arr, val0}, rr = 1/(1 - cc + cc dd);
  arr = Table[pi = data[[i]] recipN; Gamma[dd + 1, 1 - cc Log[pi]],
   {i, imax + 1}];
  val0 = rr (E Total[arr] - cc); If[iswitch == 1, val0 - cc, val0]]
```

with formal arguments `c, d, data, recipN, imax, iswitch`, where the first two ones are parameters specific for the (c, d)-entropy, `data` is to be substituted by an array containing probability distribution P, `recipN, imax` should be $1/n, n-1$, in case when n discrete state of system are represented by bins $\{0, 1, \ldots, n-1\}$, and `iswitch` is a flag taking either value 0 for original version of (c, d)-entropy, or 1 for the modified version, as given in (11).

bins:	0	1	2	3	4	5	6	7	8	9
i	1	2	3	4	5	6	7	8	9	10
p_i	.094	.092	.111	.116	.104	.110	.094	0.106	0.088	0.085

Table 2 Sample uniform distribution $U_{\rm s}$ generated by RandomInteger

In Figure 3, we show an iso-value curve $\lambda_1(c, d) = 0$, which provides a substantial generalization of Eq. (12)

$$S^*_{c,d}(U_{\rm s}) = S_{\rm BGS}(U_{\rm s}),\ \forall (c,d) \in \big\{(u,v) \in [0.1,5]^2 \,|\, \lambda_1(u,v) = S^*_{u,v}(U_{\rm s}) - s_u = 0, s_u = S_{\rm BGS}(U_{\rm s})\big\}. \quad (16)$$

In Figure 4, we show 2-D iso-value manifold $\lambda_2(c, d, q) = 0$ in 3-D, which is defined implicitly by

$$S^*_{c,d}(U_{\rm s}) = S_q(U_{\rm s}), \quad \forall (c,d,q) \in \big\{(u,v,w) \in [0.1,5]^3 \,|\, \lambda_2(u,v,w) = S^*_{u,v}(U_{\rm s}) - S_w(U_{\rm s}) = 0\big\}. \quad (17)$$

4 Conclusions

In the paper, we discussed briefly information background of operational complexity measurement of supplier-customer systems. As the operational complexity measuring procedure is based upon entropy, we focused our attention on various generalization of Shannon entropy. In sequel, we discussed Renyi, Tsallis, and Masi entropies, generalized (c, d) entropy, and multi-dimensional entropic measure. The iso-value curve and 2-D manifold, which we investigated numerically, are new results. All entropies discussed represent good platform for research on construction of effective manager purpose-oriented operational complexity measures of supplier-consumer systems.

References

[1] Filiz, I.: An entropy-based approach for measuring complexity in supply chains, *Int. J. Prod. Res.* **48** (2010), 3681–3696.

[2] Hanel, R., and Thurner, S.: Generalized (c, d)-entropy and aging random walks, *Entropy* **15**(12), (2013), doi:10.3390/e15125324, 5324–5337.

[3] Lukáš, L., and Plevný, M.: Using entropy for quantitative measurement of operational complexity of supplier-customer system: case studies, *Central European Journal of Operations Research* **24**(2) (2016), 371–387.

[4] Lukáš, L., and Hofman, J.: Operational Complexity of Supplier-Customer Systems Measured by Entropy – Case Studies, *Entropy* **18**(4), 137 (2016), doi:10.3390/e180400137, 1–17.

[5] Masi, M.: A step beyond Tsallis and Renyi entropies, *Phys. Lett. A* **338** (2005), 217–224.

[6] Maasoumi, E., and Zandvakili, S.: Generalized entropy measures of mobility for different sexes and income levels, *J. of Econometrics* **43** (1990), 121–133.

[7] Porochorov, J.B., and Rozanov, J.A.: *Teorije verojatnostej* (*Theory of probability*, in russian). Nauka, Moscow, 1967.

[8] Remuzgo, L., Trueba, C., and Sarabia, J.M.: Evolution of the global inequality in greenhouse gases emissions using multidimensional generalized entropy measure, *Physica A* **444** (2016), 146–157.

[9] Shang, D., Mengjia, X., and Shang, P.: Generalized sample entropy analysis for traffic signals based on similarity measure, *Physica A* **474** (2017), 1–7.

[10] Sivadasan, S., Efstathiou, J., Calinescu, A., and Huaccho Huatuco, L.: Advances on measuring the operational complexity of supplier-customer systems, *European J. of Oper. Rers.* **171** (2006), 208–226.

Data Envelopment Analysis of the Renewable Energy Sources

Jana Sekničková[1]

Abstract. The most of energy demand is covered by the usage of fossil fuels at the present time and so these sources are relatively quickly mined. That is why the investments into renewable energy sources are higher and higher all the time. There are five real possibilities of renewable energy sources in the Czech Republic – wind power stations, photovoltaic power stations, geothermal power stations, hydroelectric power stations and biomass energy. Although from the multicriteria point of view the photovoltaic power stations are not well evaluated the low investment costs and easy accessibility of suitable lands are strong points of photovoltaic energy and so the number of newly built photovoltaic power stations is huge. In this paper the renewable energy sources are analyzed from multicriteria point of view, especially by using data envelopment analysis because photovoltaic power stations are typical production units. The analysis is based on real data of the photovoltaic power stations built in the Czech Republic.

Keywords: Renewable energy sources, DEA models, Photovoltaic power station.

JEL Classification: C44, C67
AMS Classification: 90B30, 90B50, 90C29, 91B06, 91B38

1 Introduction

The Czech Republic as a member of the European Union is bounded by the Directive 2009/28/EC of the European Parliament on the promotion of the use of energy from renewable sources. For the Czech Republic this directive implies 13% of the overall share of energy from renewable sources before 2020. In addition the Czech Republic tends to 15% share in 2030 and 30% share of energy from renewable sources in 2050 in accordance to the State energy conception. There are five real possibilities of renewable energy sources (RES) building-up in the Czech Republic. The first ones are wind power stations for wind energy production. The second possibility is a photovoltaic power station based on obtaining of sun energy. The geothermal power station deals with thermal energy of the Earth. Places with large water areas or swift rivers are suitable for hydroelectric power stations. The last but not least possibility is to gain biomass energy from biomass. There exist several other alternative energy sources but such as energy of sea waves are not reachable in the Czech conditions.

Each of the mentioned energy sources has its advantages and also disadvantages. In [7] and [8] authors evaluate these RES from the multicriteria point of view. It is a complex decision making problem with a small number of alternatives (five in this case) and many decision criteria. For the evaluation TESES (technical, economic, social, ecological and strategic) classification of criteria was used. The technical criteria describe the technical obstacles of single technologies, technical parameters of produced electricity and if the technology enables an efficient usage of produced warm, coefficient of hazard in the building-up season etc. Economic criteria describe the project economy on the level of assessment from the investor's point of view that is conclusive for the project realization or rejection. Social criteria describe the social project implications and their benefit for solving of social and socio-economic problems of the regions and states. Ecological criteria describe the particular benefits of single technology for environment components. And strategic criteria describe the long-term project effect for power industry in the Czech Republic and for the situation in further spheres of the national economy – for example agriculture. There are used five methods of multicriteria evaluation of alternatives and geothermal power was the winner with respect to all five methods in mentioned papers. The second one was biomass power, the third wind power, the fourth and the fifth places were for photovoltaic and hydroelectric powers.

These results were very surprising because in 2010 almost 13 thousands photovoltaic power stations were installed in the Czech Republic, in comparison with about one hundred hydroelectric power stations, about 50 wind power stations with approximately 150 turbines, 5 biomass power stations and only few very small geothermal power stations. The articles about the first real geothermal power station in the Czech Republic was published in that year. The conclusion was that in the Czech Republic inefficient renewable energy sources were used often and efficient power stations were used too little.

[1] University of Economics Prague, Department of Econometrics, W. Churchill Sq. 4, Prague 3, Czech Republic, e-mail: jana.seknickova@vse.cz

The same problem was evaluated from the Data Envelopment Analysis (DEA) point of view in [9] one year later. The model with only two inputs (investment costs of project and time of recovery) and two outputs (for the first case: internal rate of return and recovery of investments, and for the second case: annual utilization of installed output and net profit) was used. Biomass power and geothermal power were found efficient with respect to all DEA models and photovoltaic power was efficient in the case of variable return to scale. Wind and hydroelectric powers were inefficient with respect to all DEA models.

In this paper only photovoltaic power stations will be analyzed due to answer the basic question: *Are installed photovoltaic power stations efficient or not?* The null hypothesis assumes that only few of analyzed photovoltaic power stations are efficient and the most of them are inefficient. In the case that this null hypothesis will be reject the answer to this basic question can help to explain why the number of installed photovoltaic power stations is so huge in comparison to the other renewable energy sources.

2 DEA point of view

Each analyzed photovoltaic power station (alternative) is evaluated from the multicriteria point of view by each investor. Evaluation of investments into RES is a multiple criteria decision making problem with the evaluated production units as alternatives (energy sources) and with the multiple inputs as criteria to be minimized (investments, recovery, etc.) and multiple outputs as criteria to be maximized (expected output, profit, etc.) in the typical case. This problem can be solved by several modeling techniques. Among them are standard multiple criteria decision making (MCDM) methods used in [7] and [8], data envelopment analysis models used in [6] and others (for example using methods of mixed integer programming at optimization of the production used in [11]).

Each photovoltaic power station is production unit and DEA is suitable tool for evaluation of efficiency of production units. The term efficiency is defined as the ratio of multiple effects (outputs) produced by the evaluated units on the one hand and multiple resources (inputs) that are spent during the transformation process of inputs into the outputs on the other hand. Typically, the number of inputs and outputs is higher than one. That is why the question is how to compare multiple outputs and multiple inputs.

First DEA models were formulated in 1978 but their main expansion starts in the last ten years of the 20th century. Application of DEA models in the field of renewable energy sources started in last ten years and now we can found papers using DEA models for evaluation of efficiency of different renewable energy sources. For example the problem of the North Africa is described in [2], the comparison of the Germany and the United States is in [13] but we didn't find similar analysis for the Czech Republic.

2.1 Decision units

For this paper the list of 28 real Czech photovoltaic power stations as decision units is given. The used data are from the real projects realized in 2008 – 2016 and they are gained from the non-profit-making organization focused on the renewable energy sources in the Czech Republic.

Note that DEA models are suitable for evaluation of a large number of production units with respect to a relative small number of inputs and outputs. It is assumed the number of 28 units is large enough.

2.2 Criteria – inputs and outputs

For DEA models we divided criteria into two groups – inputs and outputs. List of **inputs** contains six minimizing criteria and it involves:

- time of recovery in years,
- investment costs of project – capital expenditures related with building, complexity of realization, time of realization and technical complexity in thousands CZK,
- total fixed costs in thousands CZK per year,
- total operating costs in CZK per year,
- total energy consumption for operation in kWh per year,
- volume of suitable area in m^2.

List of **outputs** contains eight maximizing criteria and it involves:

- annual utilization of installed output – ratio of real energy production per year to maximal theoretical energy production in %,
- expected lifetime of power station – how long it will be able to produce energy in years,
- the net present value – NPV in thousands CZK,
- internal rate of return – IRR in %,

- decreasing of carbon dioxide – evaluated by amount of saved fugitive emissions in comparison with pitcoal in tons per year,
- decreasing of other polluting components – evaluated by amount of saved fugitive emissions in comparison with pitcoal in tons per year,
- effectivity of photovoltaic panel in %,
- revenues from energy sales in thousands CZK per year.

Statistical characteristics for complete data set of 28 units and 14 criteria are included in Table 3 in Appendix.

3 DEA Method

3.1 Data envelopment analysis

Data envelopment analysis (DEA) is a set of non-parametric techniques based on solving of linear programming problems for evaluation of efficiency of the set of homogenous units. The efficient unit consumes minimal amount of inputs and produces maximal amount of outputs. The basic idea of DEA models, developed by Charnes et al. [3, 4], consists in estimation of so-called efficient frontier, and projects all decision making units (DMUs) onto this frontier. If a DMU lies on the frontier, it is referred to as an efficient unit, otherwise inefficient. All efficient DMUs are considered equally "good" [5]. DEA also provides efficiency scores and reference units for inefficient DMUs. Reference units are hypothetical units on the efficient frontier, which can be regarded as target units for inefficient units. A virtual reference unit is traditionally found in DEA by projecting the inefficient DMU radially onto the efficient frontier. The advantage of DEA models is a fact that efficiency evaluation is based on the data available without taking into account the decision-maker's preferences.

Note also that DEA models are linear problems those can be solved classically for example by simplex method. As is written below the alternative possibility for solution is to solve dual problem [14]. The review and detailed information about DEA models can be found in [1] and [10].

3.2 Mathematical model

The mathematical model of DEA considers r DMUs $U_1, U_2, \dots, U_r$ with m inputs ($i = 1,2, \dots, m$) and n outputs ($j = 1,2, \dots, n$). The vector of input values of DMU k ($k = 1,2, \dots, r$) is denoted as $\mathbf{x}_k = (x_{1k}, x_{2k}, \dots, x_{mk})^{\mathrm{T}}$ and matrix of all input values for all DMUs is denoted as $\mathbf{X} = \{x_{ik}, i = 1,2, \dots, m, k = 1,2, \dots, r\}$. Similarly, the vector of output values of DMU k is denoted as $\mathbf{y}_k = (y_{1k}, y_{2k}, \dots, y_{nk})^{\mathrm{T}}$ and matrix of all output values for all DMUs is denoted as $\mathbf{Y} = \{y_{jk}, j = 1,2, \dots, n, k = 1,2, \dots, r\}$. The relative technical efficiency of given DMU q can be generally expressed as ratio of weighted sum of outputs and weighted sum of inputs

$$TE_q = \frac{\sum_{j=1}^{n} u_j y_{jq}}{\sum_{i=1}^{m} v_i x_{iq}}, \tag{1}$$

where v_i, $i = 1, 2, ..., m$ is a weight for i-th input and u_j, $j = 1, 2, ..., n$ is a weight for j-th output.

DEA models can be oriented to inputs or outputs. In this study fixed level of outputs (as we theoretically can change e.g. the available area or amount of investments but we are unable to influence the amount of produced CO_2 or expected durability life) is assumed and so input oriented models are used. In such case it is suitable to minimize inputs with respect to given outputs. Such model has the following form

maximize

$$z = \mathbf{u}^{\mathrm{T}} \mathbf{y}_q,$$

subject to

$$\begin{aligned} \mathbf{v}^{\mathrm{T}} \mathbf{x}_q &= 1, \\ \mathbf{u}^{\mathrm{T}} \mathbf{Y} - \mathbf{v}^{\mathrm{T}} \mathbf{X} &\leq 0, \\ \mathbf{u} &\geq \varepsilon, \\ \mathbf{v} &\geq \varepsilon. \end{aligned} \tag{2}$$

Model (2) is called primary CCR (Charnes, Cooper, Rhodes) input oriented model (CCR-I) and it assumes constant return to scale (CRS).

For practical solution dual CCR-I model can be used. It has form

minimize

$$f = \theta + \varepsilon(\mathbf{e}^{\mathrm{T}}\mathbf{s}^{+} + \mathbf{e}^{\mathrm{T}}\mathbf{s}^{-}),$$

subject to

$$\begin{aligned} \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^{+} &= \mathbf{y}_q, \\ \mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^{-} &= \theta\mathbf{x}_q, \\ \boldsymbol{\lambda}, \mathbf{s}^{+}, \mathbf{s}^{-} &\geq \mathbf{0}, \end{aligned} \tag{3}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_r)^{\mathrm{T}}$ are weights for DMUs, $\mathbf{s}^{+}$ and $\mathbf{s}^{-}$ are vectors of slack variables for inputs and outputs, $\mathbf{e}^{\mathrm{T}} = (1, 1, \dots, 1)$ a $\varepsilon > 0$ is a number smaller than any positive real number.

The evaluated q-th DMU is efficient if and only if optimal value of $\theta = 1$ and all slack variables ($\mathbf{s}^{+}$ and $\mathbf{s}^{-}$) in the optimal solution are zeros.

The models (2) and (3) are used if constant return to scale (CRS) is assumed. In the case of variable return to scale (VRS) we work with BCC (Banker, Charnes, Cooper) model and the dual model (3) must be extended about constraint $\mathbf{e}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{1}$ [12]. The output oriented models have similar mathematical model as (3), with conditions $\mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^{+} = \theta\mathbf{y}_q$, $\mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^{-} = \mathbf{x}_q$ and with maximal value of objective function.

4 Results

The problem of photovoltaic power stations was analyzed from DEA point of view. For computation optimization software Lingo 11.0 was used. As it is not easy to set that the photovoltaic power stations have constant return to scale (CRS) or variable return to scale (VRS), CCR and also BCC models were calculated. In the set of 28 DMUs (photovoltaic power stations) 20 efficient units were found in the case of CRS (CCR models) and 21 effective units in the case of VRS (BCC models). For comparison also output models were calculated. The lists of effective units are identical for input oriented models as well as for output oriented models. All efficient units from CCR models are efficient also in BCC models. The unit no. 2 is efficient in the case of variable return to scale but inefficient in the case of constant return to scale.

For all effective units the measure of super-efficiency was calculated and the best seven units are displayed in Table 1. With respect to super-efficiency it can concluded that unit no. 7 is the best alternative. From original data we can see that this power station has the lowest only total operating costs and volume of suitable area. The values of other criteria are not the best but they are above average.

unit no.	**7**	**8**	**9**	**12**	**18**	**22**	**26**
CCR-I	14.136	1.557	2.226	2.253	1.301	1.862	1.438
BCC-I	14.339	2.306	2.476	2.971	6.132	16.299	11.455
CCR-O	0.071	0.643	0.455	0.445	0.770	0.544	0.696
BCC-O	0.070	0.155	0.455	0.380	0.322	0.522	0.585

Table 1 DEA results for efficient units – super-efficiency

In Table 2 we can see values of theta parameters for all input and output oriented CCR and BCC models for the inefficient units. We can also see that in the set of units is only one evidently bad unit (no. 21) – this power station has absolutely highest time of recovery (14 years in comparison with 9-12 years for other units) and the lowest values of all eight output criteria. The theta parameter for this unit is about 0.73 in the case of input oriented models that means this power station has to reduce all inputs to 73%. In the case of time of recovery it means to ensure recovery of investment to 10.3 years.

unit no.	**2**	**4**	**6**	**10**	**20**	**21**	**24**	**28**
CCR-I	0.968	0.960	0.915	0.854	0.990	0.726	0.998	0.901
BCC-I	1.000	0.980	0.915	0.855	0.990	0.728	0.998	0.902
CCR-O	1.033	1.042	1.093	1.170	1.010	1.376	1.002	1.110
BCC-O	1.000	1.003	1.003	1.010	1.010	1.005	1.002	1.001

Table 2 DEA results for inefficient units – theta parameters

5 Conclusion

From the previous research it is known that there are five possibilities of renewable energy sources in the Czech. From multicriteria point of view and also from DEA point of view the photovoltaic power stations are not efficient investment. However the number of photovoltaic power stations is severalfold higher than in the case of other RES. In this paper list of 28 power stations was analyzed using DEA models and it was expected that only few projects is effective and the most of projects is inefficient. However the results are unexpected.

The most of photovoltaic power stations (more than 70%) is efficient and in the sample only one photovoltaic power station is explicitly bad. The results of remaining inefficient units depend probably on the other data not included in this research (only 14 criteria were considered and calculation with such important criteria as net profit of firm, value of subsidy etc. was not done) and these inefficient units can be efficient easily. Ten percentage decrease in inputs ensures efficiency of these power stations.

Based on real data it was concluded that the low investment costs and easy accessibility of suitable lands are strong points of photovoltaic energy but these criteria are not critical. The other criteria play also an important role and can significantly influence decision making of investor.

Acknowledgements

This paper was supported by the grant No. 16-01821S of the Czech Science Foundation.

References

[1] Allen, C., and Hall, S.: *Macroeconomic Modelling in a Changing World*. John Wiley & Sons Inc., Chichester, 1997.

[2] Boubaker, K.: Renewable energy in upper North Africa: Present versus 2025-horizon perspectives optimization using a Data Envelopment Analysis (DEA) frame-work. *Renewable Energy* **43** (2012), 364–369.

[3] Charnes A., Cooper W. W., and Rhodes E.: Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.

[4] Charnes A., Cooper W. W., and Rhodes E.: Short communication: measuring efficiency of decision making units. *European Journal of Operational Research* **3** (1979), 339.

[5] Hosseinzadeh Lotfi, F., Noora, A. A., Jahanshahloo, G. R., Jablonský, J., Mozaffari, M. R., and Gerami, J.: An MOLP based procedure for finding efficient units in DEA models, *CEJOR* **17** (2009), 1–11.

[6] Jablonsky, J., and Dlouhy, M.: Solving DEA and MCDM decision problems in spreadsheets. In: *Efficiency and Responsibility in Education*. Praha : Czech University of Life Sciences in Prague, 2010, 126–136.

[7] Jablonsky, J., and Kalcevova, J.: Multicriteria Evaluation of Investments into Renewable Energy Sources. In: *Quantitative Methods in Economics – Multiple Criteria Decision Making XV*. Bratislava : University of Economics, 2010, 74–84.

[8] Jablonsky, J., and Kalcevova, J.: Multicriteria analysis of investments into renewable energy sources. In: *Sustainable Development: Industrial Practice, Education & Research*. Bari : DIMEG Edizioni, 2010, 20–25.

[9] Kalcevova, J.: Renewable Energy Sources and DEA Models. *Mathematical Methods in Economics 2011*. Prague : PROFESSIONAL PUBLISHING, 2011, 327–332. ISBN 978-80-7431-058-4.

[10] Pedraja-Chaparro, F., Salinas-Jimenez, J., and Smith, P.: On the Role of weight restrictions in data envelopment analysis. *Journal of Productivity Analysis* **8** (1997), 215–230.

[11] Pelikan, J.: An Optimisation of the Production System – Case Study. In: *Mathematical Methods in Economics 2003*. Praha : ČZU, 2003, s. 203–208.

[12] Sevkovic, D., Halicka, M., and Brunovsky, P.: DEA analysis for a large structured bank branch network. *CEJOR* **9** (2001), 329-342.

[13] Sueyoshi, T., and Goto, M.: Photovoltaic power stations in Germany and the United States: A comparative study by data envelopment analysis, *Energy Economics* **42** (2014), 271–288.

[14] Thompson, R. G, Singleton, F., Thrall, R., and Smith, B.: Comparative site evaluations for locating a highenergy physics lab in Texas. *Interfaces* **16** (1986), 35–49.

Appendix

criterium	units	minimum	median	maximum	mean	standard deviation
Inputs (minimize)						
time of recovery	years	8	10	15	10.357	1.830
investment costs of project	1 000 CZK	2 980.000	78 199.160	360 000.000	98 694.494	84 852.588
fixed costs	1 000 CZK / year	40.900	1 066.920	5 790.059	1 435.966	1 446.927
operating costs	CZK / year	1 445.000	48 196.000	300 213.000	75 916.143	69 820.127
energy consumption	kWh / year	2 217.000	13 168.775	91 798.000	22 769.007	20 337.247
volume of suitable area	m^2	0.064	0.755	1.660	0.781	0.714
Outputs (maximize)						
utilization of output	%	0.190	0.216	0.257	0.216	0.017
expected lifetime	years	20	25	25	24.821	0.945
net present value	CZK	6.951	23 910.306	79 974.414	24 755.488	20 352.905
internal rate of return	%	0.061	0.125	0.314	0.142	0.059
carbon dioxide	ton / year	199.166	1 079.771	4 806.512	1 294.424	1 082.593
other polluting components	ton / year	0.614	3.959	14.806	4.568	3.034
effectivity of panel	%	0.051	0.140	0.178	0.132	0.028
revenues from energy sales	CZK / year	375 534.792	10 738 084.808	50 270 862.639	13 261 505.144	11 594 329.485

Table 3 Data set characteristics

Wavelet Concepts in stock prices analysis

Jaroslav Schürrer[1]

Abstract. In this paper, wavelet concepts are used to analyze stock prices and compare them with Geometric Brownian motion which is often used to model financial processes. Discrete Wavelet Transformation is used to obtain distribution of wavelet coefficients power spectrum among different frequency bands for stock data and Geometric Brownian motion. Apart from wavelet spectra, we also calculate the wavelet correlation coefficients to analyze correlation dependency on scale and wavelet coefficients variance to indicate fundamental changes about data generating mechanism. Geometric Brownian motion distribution is compared with stock data power distribution to measure the probability of effectiveness market of stock prices.

Keywords: Discrete Wavelet Transformation, Variance, Power spectrum, Wavelet correlation coefficients.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Mostly, wavelet transform is used in specific cases to resolve the particular problem. In this paper, however, we will focus on a more comprehensive view of how to use wavelet transformation in analysis of finantial time series. The goal of this article is to show concepts of wavelet analysis during stock prices analysis. We use Discrete Wavelet Transformation (DWT) to compute power spectrum at different frequency band for time series represented by one index and company shares. Next we use Maximal Overlap Discrete Wavelet Transform (MODWT) to partition the data variance by scale. We show conceptual algorithms which will guide us during optimisation process of wavelet transformation.

The article is divided into two main parts. First part summarizes Wavelet theory, Geometric Brownian motion and setup bases for assumption about normality of data. The second part first analyses financial stock data and random stock price movement represented by Geometric Brownian motion. All numerical results are based on own code written in Python with support of additional packages for scientific computing (NumPy and SciPy).

2 Wavelet Theory

Wavelets are functions that satisfy certain requirements. They integrate to zero, the function has to be well localized and also there exists easy calculation of the direct and inverse wavelet transform. The most simple and oldest one is Haar wavelet. Wavelets are used as basis functions in representing other functions. Wavelets are local in both frequency (scale via dilations) and in time (via translations). A wavelet function can be viewed as a high pass filter, which approximates a input data. The result of the wavelet function is the difference between value calculated by the wavelet function and the actual data. The scaling function calculates a smoothed version of the data, which becomes the input for the next iteration of the wavelet function. There are an infinite variety of wavelet and wavelet scaling functions. The closeness of the approximation provided by the wavelet function depends on the nature of the data.

For simplicity, we use Haar wavelet definition for the wavelet definition and scaling function, which is relatively simple to implement in any programming language. Wavelet is a function that results in a set of high frequency differences, or wavelet coefficients denoted cD_i. For times series $x[n] = [x_1, x_2, ..., x_n]$ we define the Haar wavelet equation

$$cD_i = \frac{x_i - x_{i+1}}{2} \tag{1}$$

where cD_i is the wavelet (difference) coefficient. The wavelet function is a high pass filter. A high pass filter allows the high frequency components of a signal through while suppressing the low frequency components. The scaling function produces a smoother version of the $x[n]$, which is half the size of the $x[n]$. The scaling function is a low pass filter.

[1]Masaryk Institute of Advanced Studies, Czech Technical University, Kolejni 2637/2a, Prague, Czech Republic, schurjar@cvut.cz

A low pass filter suppresses the high frequency components of a signal and allows the low frequency components through. Wavelet algorithms are recursive and the smoothed data becomes the input for the next step of the wavelet transform. The Haar wavelet scaling function is defined by following equation

$$cA_i = \frac{x_i + x_{i+1}}{2} \tag{2}$$

where cA_i is a smoothed value (approximation). Because Haar wavelet preserves energy of $x[n]$ we have to multiply wavelet and scaling coefficients with extra term $\frac{1}{\sqrt{2}}$.

2.1 Discrete Wavelet Transformation

We define Discrete Wavelet Transformation by terminology of filters mainly used in signal processing. To decompose a time series of length $N = 2J$ using DWT, the maximum number of levels in the decomposition is J. Length of signal also influences maximum number of decomposition levels which is defined by equation

$$floor(\log(datalen/(filterlen1))/log(2)) \tag{3}$$

This decomposition levels are sometimes called bands. The fast wavelet transform algorithm uses two filters h and g that characterize their interaction:

$$\begin{aligned} g[n] &= \frac{1}{\sqrt{2}}\langle \psi(x/2), \varphi(x-n)\rangle \\ h[n] &= \frac{1}{\sqrt{2}}\langle \varphi(x/2), \varphi(x-n)\rangle \end{aligned} \tag{4}$$

The simplest filters are the Haar filters with $h = [1,1]/\sqrt{2}$ and $g = [-1,1]/\sqrt{2}$. This definition assumes that signal has even number of data points. In numerical computation, we can use several techniques called signal extension modes:

- zero-padding - signal is extended by adding zero samples
- constant-padding - border values are replicated
- symmetric-padding - signal is extended by mirroring samples
- periodic-padding - signal is treated as a periodic one

2.2 Maximal Overlap Discrete Wavelet Transform

The Maximal Overlap Discrete Wavelet Transform is an undecimated wavelet transform over dyadic scales, which is frequently used with financial data. The MODWT differs from the DWT in that it is a highly redundant, nonorthogonal transform. The MODWT retains downsampled values at each level of the decomposition that would be otherwise discarded by the DWT. The MODWT is well defined for all sample sizes N opposite to DWT which requires N to be a multiple of 2^J for complete decomposition of J levels. One nice feature of the MODWT for time series analysis is that it partitions the data variance by scale.

2.3 Pyramidal Algorithm

Wavelets are used to decompose a time series to multiple resolutions. Each resolution reflects a different frequency. The wavelet technique takes averages and differences of a signal, breaking the signal down into spectrum. Usually the wavelet algorithms work on time series a power of two values. Each step of the wavelet transform produces two sets of values: a set of averages (cA coefficients) and a set of differences (cD coefficients which are sometimes referred to as wavelet coefficients). Each step produces a set of averages and coefficients that is half the size of the input data.

The average and difference of the time series is made across a window of values. Most wavelet algorithms calculate each new average and difference by shifting this window over the input data. At every next step of the calculation the previous set of averages is used (also shifting the window by two elements). This tree structured recursive algorithm is referred to as a Pyramidal algorithm.

2.4 Power Spectrum

A power spectrum can be calculated from the result of a wavelet transform. If the data set consists of N elements, where N is a power of two[2], there will be $\log_2(N)$ coefficient bands and one scaling value. The wavelet power spectrum is calculated by summing the squares of the coefficient values for each band:

$$spectrum_j = \sum_{k=0}^{2^j-1} c_{j,k}^2 \tag{5}$$

In the spectrum plots the average is squared as well. This results in the square of the average at $spectrum_1$ and the square of the first coefficient band at $spectrum_1$. Original signal energy is located at $spectrum_0$.

2.5 Artificial Time Series - White and Brown Noise

A white noise process can be described as follows:

$$r_n = \epsilon_n \tag{6}$$

where the $\epsilon_n \sim N(0, \sigma_\epsilon^2)$ are normal distributed numbers with zero mean. White noise is an AR(1) process where $\alpha = 0$ and $m = 0$. A brown noise process can be created from a white noise process as follows

$$B_n = B_{n-1} + r_n \tag{7}$$

This process could be described as an AR(1) process where $\alpha = 1$ and $m = 0$. This would imply that the variance σ^2 of this process tends to infinity and therefore also same hold for values of the wavelet spectrum. But since we use finite length time series and numerical methods to perform a wavelet analysis, the values of the wavelet spectrum will be finite. Generally it is important to always make sure that the time series under consideration has finite moments and can be described in some approximation as an AR(1) process. This is not true for brown noise and thats why we use differentiated brown noise to perform a wavelet analysis:

$$B_n = B_n - B_{n-1} \tag{8}$$

Note that a positive side effect of differentiating the time series is that the signal-noise ratio improves. For the same reasoning, we dont analyse financial price series directly but use the log-returns.

2.6 Wavelet analysis

In this subsection, we summarize conceptual algorithm for wavelet analysis. The base point of the wavelet transform is the length of the time series that affects the maximum level of decomposition J (number of bands) during DWT. From the practical point of view, we choose the level of decomposition $J_p < J$. Next, we start with Haar wavelet to compute power spectrum and compare it with different wavelets (usually Daubechies family, Symlets, Coiflets, Mexican Hat, Morlet, Biorthogonal...). We focus on differences between wavelets on the different bands. We continue with MODWT to partitions the data variance by scale. Exploring wavelet variance at scales gives us an interesting insight into scales, where we focus on the scales which account approximately for 80% of the variability in input data.

3 Data analysis

In the following, one index and one company price series will be presented and analyzed [3]. First, this is the FTSE 100 Index which is a share index of the 100 most highly capitalised companies listed on the London Stock Exchange. Second, Apple an American multinational technology company. The analysis will use a time interval starting on 1/1/2006 until present time. We do not directly analyse the original or logarithmic price series as outlined in section 2.5, but we use log-returns calculated as follows:

$$r_t = \log\left(\frac{S_t}{S_{t-1}}\right) = \log(S_t) - \log(S_{t-1}) \tag{9}$$

[2]There are several methods of signal extrapolation that are mainly used: zero-padding, constant-padding, symmetric-padding and periodic- padding

[3]Extracted data from Yahoo! Finance (freely available by remote data access)

Figure 1 show analysed data modified to according equation 9. We can see greater variability in Apple shares than in FTSE index. However, we still have to keep in mind that the chart shows the log-returns of differences.

Figure 1 Stock and index levels evolution (2804 entries)

In theoretical finance, there is a widely used assumption that stock price returns are normally distributed. This partially stems from the fact that models use random walk theory of stock price. Histogram of return in Figure 2 shows that real-time data might not be normal. Pdf exhibits faster tails than the normal distribution.

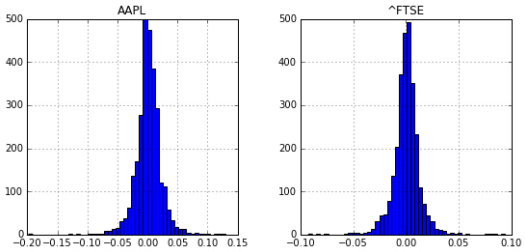


Figure 2 Stock and index histogram

Figure 3 shows the wavelet power spectrum for the result of the Haar transform applied to the signal shown in Figure 1. The point plotted at zero on the x-axis is the square of the RAW input data and next one is single scaling function value that is left after calculating the wavelet transform. The wavelet coefficient bands are plotted at x-axis points 3 to 7. The Haar wavelet provides a reasonably good approximation since both bands 4 and 5 have small values [4].

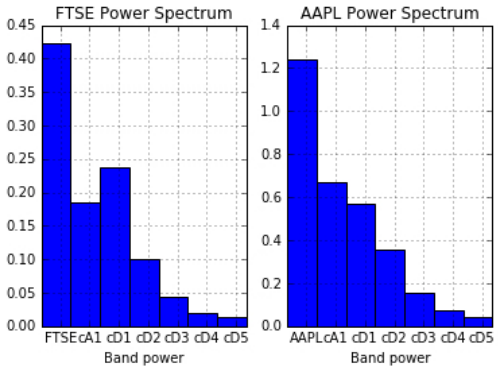


Figure 3 Stock and index power spectrum

[4]Length of signal also influences maximum number of decomposition levels which is equal in our case to 11 bands, but we can see that reasonable value is 5 bands.

We use the simplest Haar wavelet for power spectrum computation, but we can also use another wavelets. Following figure 4 show difference between power spectrum computed by Haar and DB4 wavelet. We can see that there no significant difference and that Haar wavelet is sufficient for computations.

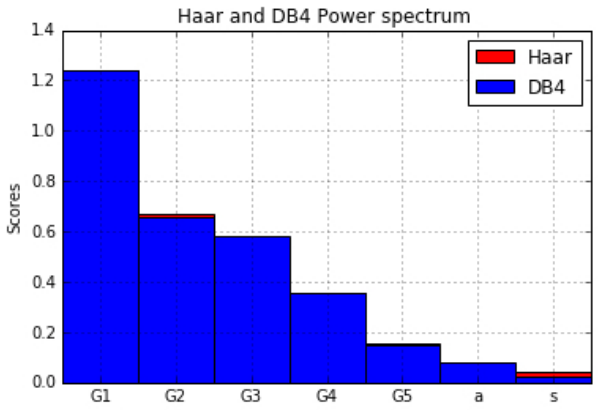


Figure 4 Haar and DB4 power spectrum of AAPL shares

Our input data include 2804 data points. We calculate MODWT down to level five. We get five arrays of wavelet coefficients variances and one of scaling coefficients variance for level 5. Figure 5 depictthe wavelet variances by scale where we ignore the scaling coefficient variance.

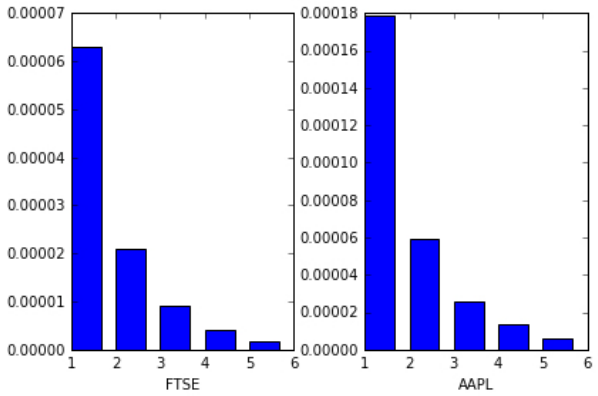


Figure 5 Wavelet variance by scales (MODWT down to level 5)

4 Conclusion

This paper presented Wavelet Analysis in power spectrum and variance by scales analysis. We demonstrate conceptual algorithm for wavelet analysis which is based on DWT and MODWT. This approach partition the data variance by scale and computes power spectrum for time series.

References

[1] Mallat, S.: *A Wavelet Tour of Signal Processing*. Academic Press, Burlington, 2008.

[2] Vidakovic, B.: *Statistical Modeling by Wavelets*. Wiley-Interscience, 111 River Street Hoboken, 1999.

[3] Ramsey, J.: The Contribution of Wavelets to the Analysis of Economic and Financial Data *Royal Society* **357** (1999), 784–786.

On estimation of bi-liner orthogonal regression parameters

Grzegorz Sitek[1]

Abstract. Orthogonal regression is one of the standard linear regression methods to correct for the effects of measurement error in predictors. A classical use of orthogonal regression occurs when two methods attempt to measure the same quality, or when two variables are related by physical laws.
Two non-parallel lines will be named as bi-lines. We considered the problem of estimation of orthogonal bi-lines function parameters. The orthogonal bi-lines function parameters are estimated using the least squares method for an implicit interdependence. In general values of parameter estimators are evaluated by means of approximation numerical method. Accuracy of the parameter estimation is analyzed.

Keywords: orthogonal bi-lines function, orthogonal linear regression, least squares method for an implicit interdependence, mixture of probability distribution.

JEL Classification: C20, C52
AMS Classification: 90C15

1 Introduction

The method of orthogonal regression has a long and distinguished history in statistics and economics. The method, which involves minimizing the perpendicular distance between the observations and the fitted line, has been viewed as superior to Ordinary Least Squares in two different contexts. First, it has often been advocated when the independent and dependent variables in a two-variable linear regression cannot be predetermined. This is because the minimizing of perpendicular distance does not depend on a specific axis Second, the method has also been extensively used when there are errors in the independent variables and, for this reason, is sometimes called the errors-in- variables model (see Fuller [3]). It has become an indispensable tool in a variety of disciplines including chemometrics, system identification, astronomy, computer vision, and econometrics. Principal Components Analysis can be used to fit a linear regression that minimizes the perpendicular distances from the data to the fitted model. In two dimensions, the orthogonal regression line therefore corresponds to the first principal component.

Antoniewicz in [1] and [2] proposed to approximate probability distribution of a one-dimensional random variable by means of two points. Generalizing this result he approximates a two-dimensional distribution by means of two lines, which he called bi-linear regression. In statistical literature this term is rather used to define specific linear model, see e.g. Gabriel in [4]. In this paper we will called Antoniewicz's model simply as bi-lines function, because in general it leads to approximation two-dimensional data spread by means of two lines.

2 The least squares method for an implicit interdependence

Two lines, none of which is parallel to the axis of the system, are described by the equation (Antoniewicz [1])

$$(y - ax - b)(y - cx - d) = 0\,. \tag{1}$$

Let *(X,Y)* be two-dimensional random variable. Antoniewicz in [1] proposed original method of approximation distribution by means of two lines, none of which is parallel to the axis of the system. Parameters a, b, c and d of the lines minimize the following function:

$$\Phi(a,\beta,\gamma,\delta) = E\left(Y - aX - c\right)^2\left(Y - bX - d\right)^2 \tag{2}$$

Based on the available data the parameters a, b, c and d will be estimated. This is equivalent to finding the straight lines that gives the best fit (representation) of the points in the scatter plot of the response versus the predictor variable. We estimate the parameters using the popular least squares method, which gives the lines that minimizes the sum of squares of the perpendicular distances from each point to the lines. The perpendicular

[1]Department of Statistics, Econometrics and Mathematics Management Faculty, University of Economics in Katowice, grzegors12@wp.pl.

distances represent the errors in the response variable. The sum of squares of these distances can then be written as follows:

$$S(a,b,c,d)=\sum_{i=1}^{n}\frac{[(y_i-a\cdot x_i-c)(y_i-b\cdot x_i-d)]^2}{(a^2+1)(b^2+1)} \tag{3}$$

The values of estimators $\hat{a}, \hat{b}, \hat{c}, \hat{d}$ which minimize S(a,b,c,d) are derived by Antoniewicz in[1]. In general, values of $\hat{a}, \hat{b}, \hat{c}, \hat{d}$ can be obtained only numerically. The least squares regression lines are given by:

$$\hat{y}_1=\hat{a}\cdot x+\hat{b},\quad \hat{y}_2=\hat{c}x+\hat{d}$$

If c=d=0 expression (3) reduces to the following

$$S(a,b)=\sum_{i=1}^{n}\frac{[(y_i-ax_i)(y_i-bx_i)]^2}{(a^2+1)(b^2+1)} \tag{4}$$

We denote $a=tg\alpha$ and $b=tg\beta$

$$S(\alpha,\beta)=\sum_{i=1}^{n}[(y_i-tg\alpha\cdot x_i)(y_i-tg\beta\cdot x_i)\cos\alpha\cos\beta]^2 \tag{5}$$

Values of $\hat{\alpha}, \hat{\beta}$, that minimize S(α,β) are given by the solution (roots) of the following nonlinear system of two equations (6).

$$\begin{cases}\sum_{i=1}^{n}[-2\cos^2\beta\cdot x_i\cdot(y_i-tg\alpha\cdot x_i)(y_i-tg\beta\cdot x_i)^2-2\cos\alpha\cos^2\beta\sin\alpha(y_i-tg\alpha\cdot x_i)^2(y_i-tg\beta\cdot x_i)^2]=0\\ \sum_{i=1}^{n}[-2\cos^2\alpha\cdot x_i\cdot(y_i-tg\alpha\cdot x_i)^2(y_i-tg\beta\cdot x_i)-2\cos^2\alpha\cos\beta\sin\beta(y_i-tg\alpha\cdot x_i)^2(y_i-tg\beta\cdot x_i)^2]=0\end{cases} \tag{6}$$

Under the assumption that $\alpha\neq\beta$, after appropriate transformations we have:

$$\sum_{i=1}^{n}[-2\sin(\alpha-\beta)\cdot(\sin\alpha\cdot x_i-\cos\alpha\cdot y_i)(\sin\beta\cdot x_i-\cos\beta\cdot y_i)\cdot(x_i^2+y_i^2)=0$$

$$a=tg\alpha=\frac{m_{22}+m_{04}-b(m_{31}+m_{13})}{m_{31}+m_{13}-b(m_{22}+m_{40})}. \tag{7}$$

where $m_{uv}=\sum_{i=1}^{n}x_i^u y_i^v$, u=0,1,2,…, v=0,1,2,…

After putting the right site of equation (6) to the first equation of system (7) we obtain the following quartic equation,

$$\begin{aligned}&b^4(m_{31}(m_{31}m_{13}-m_{22}{}^2)+b^3(m_{22}{}^2(m_{22}+m_{40})-m_{31}{}^2(m_{22}+m_{04}))+b^2(m_{31}^3m_{31}-m_{13}m_{22}(m_{22}+2m_{40})+\\&m_{31}(n_{22}^2-m_{13}^2+2m_{04}m_{22}))+b(-m_{22}^3-m_{13}^2m_{22}+m_{13}^2m_{40}-m_{04}m_{22}{}^2)+m_{13}(m_{22}{}^2-m_{31}m_{13})=0\end{aligned} \tag{8}$$

3 Bi-lines of two-dimensional distribution

Let $\mu_{uv}=E(X^uY^v)$, u=1,2,.. and v=1,2,.. be mixed moment of random variable X and Y distribution. The estimators $\hat{b}_i$, i=1,2, are a function of sample moments m_{uv}, u=0,1,2,…, v=0,1,2,…,

3.1 Bi-lines of two-dimensional normal distribution

We consider the bivariate normal distribution: N(0,0,1,1,ρ). It is well-known:

$\mu_{40} = \mu_{04} = 3, \mu_{31} = \mu_{13} = 3\rho,\ \mu_{22} = 1 + 2\rho^2$

On the basis of previously obtained results we have:

$$b_1 = -1,\ b_2 = 1, b_3 = \frac{2 + \rho^2 - \sqrt{4 - 5\rho^2 + \rho^4}}{\rho}, b_4 = \frac{2 + \rho^2 + \sqrt{4 - 5\rho^2 + \rho^4}}{\rho}, \rho \neq 0$$

Roots b_1, b_2, b_3, b_4 are critical points. One can distinguish whether a critical point is a local maximum or local minimum by using the first derivative test. We check where the derivative of the function changes from negative to positive. If $\rho > 0$

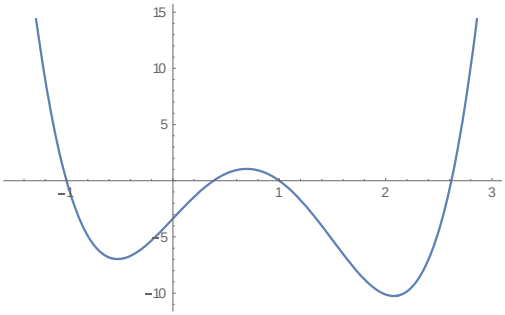


Figure 1 Graph of the polynomial (8) for ρ>0

If $\rho < 0$

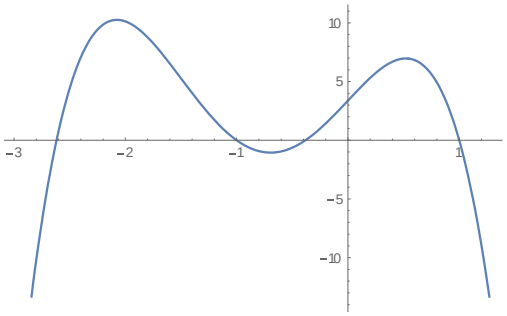


Figure 2 Graph of the polynomial (8) for ρ<0

We can see, the function S always reaches a minimum of points b_3, b_4.

We consider normal distribution N(0,0,σ_1, σ_2,ρ) In this case we have:
$\mu_{40} = 3\sigma_1^4, \mu_{04} = 3\sigma_2^4, \mu_{31} = 3\rho\sigma_1^3\sigma_2, \mu_{13} = 3\rho\sigma_2^3\sigma_1,\ \mu_{22} = \sigma_1^2\sigma_2^2(1 + 2\rho^2)$, where $\mu_{20} = \sigma_1^2$, $\mu_{02} = \sigma_2^2$.

On the basis of previously obtained results we have:

$$b_1 = \frac{\sigma_2^2 - \sigma_1^2 - \sqrt{4\rho^2\sigma_2^2\sigma_1^2 + (\sigma_2^2 - \sigma_1^2)^2}}{2\rho\sigma_2\sigma_1},\ b_2 = \frac{\sigma_2^2 - \sigma_1^2 + \sqrt{4\rho^2\sigma_2^2\sigma_1^2 + (\sigma_2^2 - \sigma_1^2)^2}}{2\rho\sigma_2\sigma_1}$$

$$-B = 6\rho\sigma_2\sigma_1^5 + 4\rho\sigma_2^3\sigma_1^3 + 8\rho^3\sigma_2^3\sigma_1^3 + 6\rho\sigma_2^5\sigma_1,\ A = 3\sigma_1^6 - 2\sigma_2^2\sigma_1^4 + 8\rho^2\sigma_2^2\sigma_1^4 - \sigma_2^4\sigma_1^2 + 4\rho^2\sigma_2^4\sigma_1^2$$

$$C = -\sigma_2^2\sigma_1^4 + 4\rho^2\sigma_2^2\sigma_1^4 - 2\sigma_2^4\sigma_1^2 + 8\rho^2\sigma_2^4\sigma_1^2 + 3\sigma_2^6,\ \Delta = B^2 - 4AC$$

$$b_3 = \frac{-B - \sqrt{\Delta}}{2A}, b_4 = \frac{-B + \sqrt{\Delta}}{2A},$$

We can see, b_2 is the classical orthogonal least squares slope estimator.

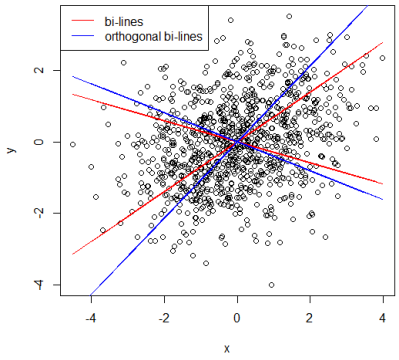


Figure 3. Bi-lines and orthogonal bi-lines of bivariate normal distribution.

3.2 Bi-lines of mixture of bivariate normal distribution

Let the distribution be defined by the following mixture of bivariate normal distribution:

pN(0,0,1,1,ρ_1)+(1-p)N(0,0,1,1,ρ_2). In this case the moments are as follows:

$$\mu_{40} = \mu_{04} = 3, \mu_{31} = \mu_{13} = 3p\rho_1 + 3(1-p)\rho_2,\ \mu_{22} = p(1+2\rho_1^2) + (1-p)(1+2\rho_2^2),$$

On the basis of previously obtained results we have:

$$b_1 = -1,\ b_2 = 1, B = 4 + 2p\rho_1^2 + 2\rho_2^2 - 2p\rho_2^2,\ A = C = 3p\rho_2 - 3\rho_2 - 3p\rho_1,\ \Delta = B^2 - 4AC$$

$$b_3 = \frac{-B - \sqrt{\Delta}}{2A}, b_4 = \frac{-B + \sqrt{\Delta}}{2A},$$

The function S reaches a minimum of points b_3, b_4 if $\rho_1 \cdot \rho_2 > 0$ and function S reaches a minimum of points b_1, b_2 if $\rho_1 \cdot \rho_2 < 0$.

4 Numerical example

This data set gives the gross national product (GNP) per capita in 1996 for various countries as well as their estimated carbon dioxide (CO2) emission per capita for the same year. The data are available in the mixtools package in R. This data frame consists of 28 countries and the following variables:

- GNP – The gross national product per capita in 1996.
- CO2 – The estimated carbon dioxide emission per capita in 1996.

As an example, we fit 2-component model to the GNP data shown in Figure 4. We use the least squares method for an implicit interdependence applying the function optim in R. The least squares orthogonal regression lines are given by:

$$\hat{CO2}_1 = -0.094 \cdot \text{GNP} + 10.71\ \ \hat{CO2}_2 = 0.634 \cdot \text{GNP} + 0.96$$

The coefficient of determination is 0.88.

The least squares classical regression lines are given by:

$\hat{CO2}_1 = -0.06 \cdot \text{GNP} + 9.51\ \ \hat{CO2}_2 = 0.83 \cdot \text{GNP} - 4.09$ The coefficient of determination is 0.93 and is higher than in the case of parameter estimation by orthogonal regression.

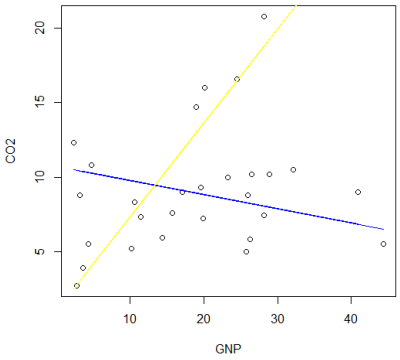


Figure 4. Orthogonal regression lines

5 Simulation analysis accuracy estimation

We calculate the parameters b_3 and b_4 based on the formulas in chapter 3.1. Next we generate n-element sample from two-dimensional normal distribution with expected values equal to zero, variances equal to one and correlation coefficient ρ. We set estimators $\hat{b}_3$ and $\hat{b}_4$ on the basis of the sample.. This operation is replicated independently N=10000 times. Finally, means of the calculated values of $\hat{b}_3$ and $\hat{b}_4$ are evaluated and denoted by a and b.

$$a=\frac{1}{N}\sum_{i=1}^{N}\hat{b}_{3,i}\quad sd_1=\sum_{i=1}^{N}(\hat{b}_{3,i}-b_3)^2\quad v_1=\frac{sd_1}{b_3}\quad b=\frac{1}{N}\sum_{i=1}^{N}\hat{b}_{4,i}\quad sd_2=\sum_{i=1}^{N}(\hat{b}_{4,i}-b_4)^2\quad v_2=\frac{sd_2}{b_4}$$

	ρ=0.4 b_3= 0.3 b_4=3.29				ρ=0.6 b_3= 0.46 b_4=2.16				ρ=0.8 b_3= 0.64 b_4=1.56			
	a (sd_1)	b (sd_2)	v_1	v_2	a (sd_1)	b (sd_2)	v_1	v_2	a (sd_1)	b (sd_2)	v_1	v_2
n=50	0.24 (0.37)	5.02 (12.48)	1.24	3.79	0.40 (0.42)	4.28 (6.54)	0.91	3.03	0.55 (0.17)	2.34 (4.41)	0.3	1.88
n=100	0.26 (0.28)	4.87 (9.83)	0.92	2.99	0.41 (0.35)	3.34 (2.46)	0.76	1.14	0.61 (0.11)	1.73 (0.57)	0.18	0.33
n=200	0.28 (0.22)	4.12 (3.49)	0.72	1.06	0.42 (0.19)	2.93 (1.41)	0.43	0.65	0.63 (0.07)	1.57 (0.20)	0,12	0.13
n=500	0.29 (0,12)	3.34 (2.79)	0.41	0.85	0.45 (0.09)	2.5 (0.76)	0.2	0.35	0.64 (0.05)	1.56 (0.12)	0.08	0.08

Table 1. Accuracy of estimation.

6 Conclusion

In this article, we considered the problem of estimation of orthogonal bi-lines function parameters by means of implicit least square method. Usually, this method needs some numerical methods because in general case it is

not possible to get exact results. In this article, we evaluate the accuracy of parameter estimation of orthogonal bi-lines function. The parameters of orthogonal bi-lines function determined on the basis of the sample are compared with the theoretical values of the parameters. Satisfactory estimation results are obtained for large samples and high correlation.

References

[1] Antoniewicz R: *Metoda najmniejszych kwadratów dla zależności niejawnych i jej zastosowanie w ekonomii.* Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 1988.
[2] Antoniewicz R.: Biproste regresji. Zeszyty Naukowe Uniwersytetu Szczecińskiego No. 318, *Prace Katedry Ekonometrii i Statystyki* **10** (2001), 9-15.
[3] Fuller, W.A.: *Measurement error models*, Wiley, New York, 1987.
[4] Gabriel K. R.: Generalised Bilinear Regression. *Biometrika* **85** (1998), 689-700.

Multiple asset portfolio with present value given as a discrete fuzzy number

Joanna Siwek[1]

Abstract. The article presents an analysis of a multiple asset portfolio with imprecise present value modelled by a discrete fuzzy number. In existing literature, the imprecision burdening a portfolio is usually modelled by given fuzzy return rates. The rates are assumed to include imprecision of information used by investor to make a financial decision, that is data which may possibly be faulty or incomplete. In the presented paper imprecision is understood as ambiguity and vagueness of information, and stems from investor's subjectivity as well as technical and technological premises. Previous research shows, that measures of ambiguity and vagueness should be calculated for discount factors rather than for return rates from a portfolio. Means of appointing the measures and some interesting conclusions about the relationship between portfolio construction and imprecision are stated throughout the paper. Finally, numerical example presenting model characteristics and exemplary calculation of portfolio fuzzy discount factor is given.

Keywords: portfolio, fuzzy number, present value, imprecision

JEL Classification: C44, C02, G10
AMS Classification: 03E72

1 Introduction

Despite that existing financial portfolio models minimize uncertainty risk, investing in effective portfolios does not guarantee a positive revenue. Even when minimizing the risk, the investor is still exposed to a kind of risk they can sense subconsciously and which affects their financial decisions. Imprecision, connected with imprecise information based on which investors makes their decision, may be a source of at least a part of this risk.

After [10], imprecision is commonly understood as ambiguity and indistinctness of information. The information ambiguity is interpreted as a lack of unambiguous distinction between the recommended alternatives. The indistinctness, on the other hand, is interpreted as a lack of distinction between a recommended and not recommended.

Imprecision risk consists of both ambiguity and indistinctness risk. An increase in ambiguity risk can result in a larger number of alternative investment recommendations. This can cause an increase in the risk of making a wrong financial decision, which *ex post* will be burdened with a loss of unexploited chances. An increase in indistinctness risk results in blurred boundaries between recommended alternatives. This, in consequence, can cause an increase in the risk of choosing a not recommended alternative.

The present state of knowledge about researched topic has its source in works of Ward [22]. He defines the present value as a discounted fuzzy cash flow. Kuchta [110] and Lesage [12] introduced fuzzy arithmetic to compute such factors as net present value. Huang [8] generalizes Ward's definition to the case of future cash flows given as a fuzzy variable. A full compendium of fuzzy portfolio analysis can be found in [5].

Piasecki [14] showed that imprecise present value and uncertain future value modelled as a variable, allow for depicting a return rate of an asset as a fuzzy probabilistic set. By generalizing this idea, in [15,16], authors researched the impact of behavioural premises and imprecision on the investor's decisions. In [17,18] a portfolio risk analysis was performed, describing the case of component assets with present value given modelled by triangular and trapezoidal fuzzy number.

The term of discrete fuzzy number, defined as a special case of fuzzy sets researched by Zadeh [23], was introduced by Voxman [21]. The arithmetic of this kind of fuzzy numbers was developed in [1,6,7]. Vicente Riera and Torrens [19,20] used the discrete fuzzy numbers to model an incomplete quantitative information.

In order to assess the financial instruments, we will use the concepts of discount factor risk measures. We choose discount factors instead of return rates because of the simplicity of measures equations while providing

[1]Poznan University of Economics, Department of Operation Research, *joanna.siwek@ue.poznan.pl*

the same information about risk. The energy measures for fuzzy sets was researched in [3,9], and in case of discrete fuzzy numbers, by Czogała, Gottwald, Pedrycz [2].

In the following article we focus on describing the imprecision risk of the portfolio rather than describing its uncertainty. This approach was decided on due to the subjective character of appointing the present value and impossibility of its verification. The present value is given subjectively by the investor and is based on their preferences, experience and other behavioral characteristics. Also, because of the discrete character of computations, as well as technical and technological premises such as delays and equipment quality, author has decided to model the imprecise present value as a discrete fuzzy number.

2 Elements of fuzzy numbers theory

A fuzzy number $\mathcal{R}$ is a fuzzy subset of $\mathbb{S}(\mathcal{R}) \subset \mathbb{R}$, described by its membership function $\mu_{\mathcal{R}}: \mathbb{S}(\mathcal{R}) \to [0,1]$ with following conditions:

$$\exists_{x \in \mathbb{S}(\mathcal{R})}: \quad \mu_{\mathcal{R}}(x) = 1 \tag{1}$$

$$\forall_{(x,y,z) \in \mathbb{S}(\mathcal{R})^3}: x \le y \le z \Longrightarrow \mu_{\mathcal{R}}(y) \ge \min\{\mu_{\mathcal{R}}(x), \mu_{\mathcal{R}}(z)\} \tag{2}$$

where $\mathbb{S}(\mathcal{R}) = \{x \in \mathbb{R}: \mu_{\mathcal{R}}(x) > 0\}$ is called a support of the fuzzy number $\mathcal{R}$. The definition was slightly altered to include discrete fuzzy numbers, which were introduced after the original definition [4].

Let us consider a pair $(\mathcal{Q}, \mathcal{R})$ of generalized fuzzy numbers described respectively by their membership functions $\mu_{\mathcal{Q}}: \mathbb{S}(\mathcal{Q}) \to [0,1]$ and $\mu_{\mathcal{R}}: \mathbb{S}(\mathcal{R}) \to [0,1]$. In agreement with the Zadeh's Extension Principle, the sum $\mathcal{Q} \oplus \mathcal{R}$ is also a fuzzy number with membership function $\mu_{\mathcal{Q}+\mathcal{R}}: \mathbb{S}(\mathcal{Q}+\mathcal{R}) \to [0,1]$ where:

$$\mathbb{S}(\mathcal{Q}+\mathcal{R}) = \{z \in \mathbb{R}: \exists_{(x,y) \in \mathbb{S}(\mathcal{Q}) \times \mathbb{S}(\mathcal{R})}: \ z = x + y\}, \tag{3}$$

$$\mu_{\mathcal{Q}+\mathcal{R}}(z) = \max\{\min\{\mu_{\mathcal{Q}}(x), \mu_{\mathcal{R}}(y)\}\}, (x,y) \in \mathbb{S}(\mathcal{Q}) \times \mathbb{S}(\mathcal{R}), z = x + y. \tag{4}$$

A discrete fuzzy number, first defined by Voxman [21], is a fuzzy number $\mathcal{R}$ with support $\mathbb{S}(\mathcal{R}) = \{x_1^{\mathcal{R}}, x_2^{\mathcal{R}}, \dots, x_{n_{\mathcal{R}}}^{\mathcal{R}}\}$. By a discretization net of a discrete fuzzy number $\mathcal{L}$ we understand the set

$$Net(\mathcal{L})) = \{x_1^{\mathcal{L}}, x_2^{\mathcal{L}}, \dots, x_{n_{\mathcal{L}}}^{\mathcal{L}}\}, \tag{5}$$

where $n_{\mathcal{L}} \in \mathbb{N}$.

Trapezoidal discrete fuzzy number, as defined in [21], is a quadruplet $DT(a,b,c,d)$ with a support $\mathbb{S}(DT(a,b,c,d)) \subset Net(DT(a,b,c,d))$, and membership function $\mu: \mathbb{S}(DT(a,b,c,d)) \to [0,1]$:

$$\mu_{DT}(x|a,b,c,d) = \begin{cases} \dfrac{x-a}{b-a}, & for \quad a \le x \le b \\ 1, & for \quad b < x < c \\ \dfrac{x-d}{c-d}, & for \quad c \le x \le d \end{cases}. \tag{6}$$

A triangular discrete fuzzy number is a special case of the trapezoidal discrete fuzzy number when $b = c$.

2.1. Discrete model

In the researched case we assume that the present value of an arbitrary asset A_i, denoted PV_i is estimated by a trapezoidal fuzzy number $T\left(\check{C}^i{}_{min}; \check{C}_*{}^i; \check{C}^{*i}; \check{C}^i{}_{max}\right)$, determined by membership function $\mu_i \in [0,1]^{\mathbb{R}}$ described by (6). Parameters of the trapezoidal fuzzy number $T\left(\check{C}^i{}_{min}; \check{C}_*{}^i; \check{C}^{*i}; \check{C}^i{}_{max}\right)$ are given as follows:

- $\check{C}^i$ is the market price,
- $\check{C}^i{}_{min} \in \left]0; \check{C}\right]$ is the maximal lower bound of PV_i,
- $\check{C}^i{}_{max} \left[\check{C}; +\infty\right[$ is the minimal upper bound of PV_i,
- $\check{C}_*{}^i \in \left[\check{C}_{min}; \check{C}\right]$ is the minimal upper assessment of prices visibly lower than the market price $\check{C}^i$,
- $\check{C}^{*i} \in \left[\check{C}; \check{C}_{max}\right]$ is the maximal lower assessment of prices visibly higher than the market price $\check{C}^i$.

The support of PV_i is

$$\mathbb{S}(PV_i) = \{x_1, x_2, \dots, x_m\} \subset Net(PV) = \{x_j \in \mathbb{R}:\ x_j = 0{,}01 \cdot j, j \in \mathbb{N}\}. \tag{7}$$

where the 0.01 unit is connected with the accuracy of computations and corresponding to 1/100 of a monetary unit. A simple return rate from an asset is defined as

$$r_t = \frac{V_t - V_0}{V_0} \tag{8}$$

where V_t is the future value of an asset and V_0 is its present value. For a given moment $t \in T$, future value of an asset A_i is described by a variable $\tilde{V}_i^t: \Omega \to \mathbb{R}$

$$\tilde{V}_t^i(\omega) = \check{C}_i \cdot \left(1 + r_t^i(\omega)\right) \tag{9}$$

Under the condition of simple return rates $r_t \in \mathbb{R}$, the membership function $\rho_i: \mathbb{R} \to [0,1]$ of a fuzzy return rate R_i from an asset with PV_i is given by the following formula:

$$\rho_i(r, \omega) = \max\left(\mu_i(x_j): r = \frac{\check{C}_i \cdot \left(1 + r_t^i(\omega)\right) - x_j}{x_j}\right) = \mu\left(\frac{\check{C}_i \cdot \left(1 + r_t^i(\omega)\right)}{1 + r}\right) \tag{10}$$

where $x_j \in \mathbb{S}(R_i)$. The support of the return rate R_i is given by the set

$$\mathbb{S}(R_i) = \left\{r \in \mathbb{R}:\ r = \frac{100 \cdot \check{C}_i \cdot \left(1 + r_t^i(\omega)\right) - j}{j}, 100 \cdot \check{C}^i{}_{min} < j < 100 \cdot \check{C}^i{}_{max}, j \in \mathbb{N}\right\} \tag{11}$$

The membership function $\rho_i: \mathbb{R} \to [0,1]$ of a fuzzy expected return rate from an asset is given by the formula

$$\rho_i(r) = \begin{cases} \dfrac{\frac{\check{C}^i \cdot (1+\bar{r}_i)}{1+r} - \check{C}^i{}_{min}}{\check{C}_*{}^i - \check{C}^i{}_{min}} & \text{dla} \quad \check{C}^i{}_{min} \le \dfrac{\check{C}^i \cdot (1 + \bar{r}_i)}{1 + r} \le \check{C}_*{}^i, \\ 1 & \text{dla} \quad \check{C}_*{}^i < \dfrac{\check{C}^i \cdot (1 + \bar{r}_i)}{1 + r} < \check{C}^{*i}, \\ \dfrac{\frac{\check{C}^i \cdot (1+\bar{r}_i)}{1+r} - \check{C}^i{}_{max}}{\check{C}^{*i} - \check{C}^i{}_{max}} & \text{dla} \quad \check{C}^{*i} \le \dfrac{\check{C}^i \cdot (1 + \bar{r}_i)}{1 + r} \le \check{C}^i{}_{max}, \\ 0 & \text{dla} \quad \check{C}^i{}_{min} \ge \dfrac{\check{C}^i \cdot (1 + \bar{r}_i)}{1 + r}, \dfrac{\check{C}^i \cdot (1 + \bar{r}_i)}{1 + r} \ge \check{C}^i{}_{max}. \end{cases} \tag{12}$$

A discount factor v, appointed based on a simple return rate r, can be calculated from

$$v = \frac{1}{1 + r}. \tag{13}$$

A fuzzy expected discount factor D_i calculated from aa fuzzy expected return rate R_i is given by membership function

$$\eta_i(v) = \begin{cases} \dfrac{\check{C}^i v - \bar{v}_i \cdot \check{C}^i{}_{min}}{\bar{v}_i \cdot \check{C}_*{}^i - \bar{v}_i \cdot \check{C}^i{}_{min}} & \text{dla} \quad \dfrac{\bar{v}_i}{\check{C}^i} \cdot \check{C}^i{}_{min} \le v \le \dfrac{\bar{v}_i}{\check{C}^i} \cdot \check{C}_*{}^i, \\ 1 & \text{dla} \quad \dfrac{\bar{v}_i}{\check{C}^i} \cdot \check{C}_*{}^i < v < \dfrac{\bar{v}_i}{\check{C}^i} \cdot \check{C}^{*i}, \\ \dfrac{\check{C}^i v - \bar{v}_i \cdot \check{C}^i{}_{max}}{\bar{v}_i \cdot \check{C}^{*i} - \bar{v}_i \cdot \check{C}^i{}_{max}} & \text{dla} \quad \dfrac{\bar{v}_i}{\check{C}^i} \cdot \check{C}^{*i} < v \le \dfrac{\bar{v}_i}{\check{C}^i} \cdot \check{C}^i{}_{max}, \\ 1 & \text{dla} \quad \dfrac{\bar{v}_i}{\check{C}^i} \cdot \check{C}^i{}_{max} < v, \qquad v < \dfrac{\bar{v}_i}{\check{C}^i} \cdot \check{C}^i{}_{min} \end{cases} \tag{14}$$

In this article, we will use energy measure for measuring ambiguity of a discount factor from given asset. Similarly as in Czogała, Gottwald, Pedrycz [2], for an arbitrary asset with present value given as a discrete fuzzy number, we can calculate the energy measure as follows

$$d(D_i) = \sum_{v \in \mathbb{S}(D_i)} \eta_i(v) = \sum_{r \in \mathbb{S}(R_i)} \rho_i(r) = \sum_{r \in \mathbb{S}(R_i)} \mu_i\left(\frac{\check{C}_i(1+\bar{r}_i)}{1+r}\right). \tag{15}$$

The entropy measure measures the indistinctness of a discount factor from an asset. For an arbitrary asset A_i with a present value given as a discrete fuzzy number, it can be calculated as follows

$$e(D_i) = \frac{\sum_{v \in \mathbb{S}(D_i)} \min\{\eta_i(v), 1-\eta_i(v)\}}{\sum_{v \in \mathbb{S}(D_i)} \max\{\eta_i(v), 1-\eta_i(v)\}}. \tag{16}$$

3 Portfolio structure and characteristics

The proposed model is based on a model introduced in [14] for an imprecise return rate and corresponds to Markowitz's model [13] with following modifications:

- we assume simple return rates,
- future value is modelled by a variable,
- present value is modelled by a discrete fuzzy number.

Let us take $n \in \mathbb{N}$ arbitrary financial assets $A_i, i = 1,2,\dots,n$, with present values PV_i given as discrete fuzzy numbers represented by their membership functions $\mu_i: \mathbb{R} \to [0,1]$. Future values of the assets modeled by $\tilde{V}^i: \Omega \to \mathbb{R}$ and market prices appointed for the moment of calculating present values are $\check{C}_i$. We assume, that simple return rates from the assets have normal distribution $N(\bar{r}_i, \sigma_i)$. From those assets we create a portfolio π. Then, according to (3) and (4), a fuzzy set PV_π referring to the present value of the portfolio takes the form of

$$PV_\pi = \sum\nolimits_{i=1}^{n} PV_i \tag{17}$$

where $x \in \mathbb{S}(PV_\pi)$ and the support of PV_π is given by the set

$$\mathbb{S}(PV_\pi) = \left\{x_j = 0{,}01 \cdot j : 100 \cdot \sum_{i=1}^{n} \check{C}^i{}_{min} < j < 100 \cdot \sum_{i=1}^{n} \check{C}^i{}_{max}, j \in \mathbb{N}\right\} \subset Net(PV). \tag{18}$$

The present value defined above for a portfolio fulfils the conditions of a discrete fuzzy number.

Under the assumption of simple return rates, the membership function of a fuzzy expected return rate from a portfolio is given as follows

$$\rho(r) = \max\left(\mu(x_i): r = \frac{\check{C}(1+\bar{r}) - x_i}{x_i}\right) = \mu\left(\frac{\check{C}(1+\bar{r})}{1+r}\right) \tag{19}$$

where $x_i \in \mathbb{S}(PV_\pi)$. The support of a fuzzy number corresponding to the expected return rate R_π of a portfolio is given by

$$\mathbb{S}(R_\pi) = \left\{r \in \mathbb{R}: r = \frac{100 \cdot \check{C}(1+\bar{r}) - j}{j}, 100 \cdot \sum_{i=1}^{n} \check{C}^i{}_{min} < j < 100 \cdot \sum_{i=1}^{n} \check{C}^i{}_{max}, j \in \mathbb{N}\right\} \tag{20}$$

Thus, the expected return rate of a portfolio defined above is a discrete fuzzy number.

Energy measure of the portfolio can be then calculated as

$$d(D_\pi) = \sum_{v \in \mathbb{S}(D_\pi)} \eta(v) = \sum_{r \in \mathbb{S}(R_\pi)} \rho(r) = \sum_{r \in \mathbb{S}(R_\pi)} \mu\left(\frac{\check{C}(1+\bar{r})}{1+r}\right) \tag{21}$$

Similarly, entropy measure for a portfolio can be calculated from

$$e(D_\pi) = \frac{\sum_{v\in\mathbb{S}(D_\pi)} \min\{\eta(v), 1-\eta(v)\}}{\sum_{v\in\mathbb{S}(D_\pi)} \max\{\eta(v), 1-\eta(v)\}}. \tag{22}$$

4 Case study

Simulations of portfolio behaviour were performed in Matlab software, researching the case of four financial assets, that is shares of polish market companies, that is BZWBK, ENEA, KGHM and ORANGE. Simple return rates were calculated based on historical data prom the period of 26.10.2016 – 25.11.2016, and parameters of present values were given subjectively by an expert, corresponding to the market price appointed for 25.11.2016.We have the following input data:

asset\parameter	$\breve{C}^i_{min}$	$\breve{C}^i_*$	$\breve{C}^{*i}$	$\breve{C}^i_{max}$	$\breve{C}_i$	$\bar{r}_i$
A1	310	313	313	315	313	-0,0046
A2	10.1	10.25	10.25	10.5	10.25	-0,0031
A3	70.5	71	73	75	72	0,0109
A4	5.5	5.55	5.65	5.7	5.6	-0,0034

Table 1 Model parameters.

The membership functions of values first asset and portfolio, their return rate of the portfolio as well as its fuzzy discount factor are presented in the pictures below.

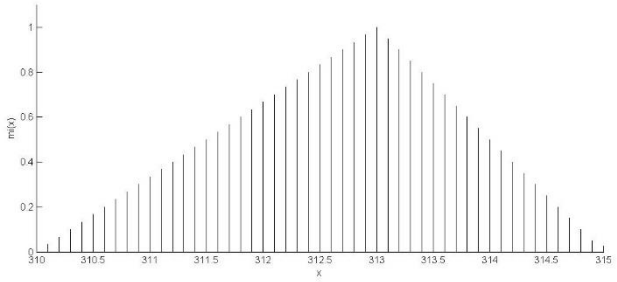


Figure 1 Exemplary present value of assets $A_i, i = 1,2,3,4$.

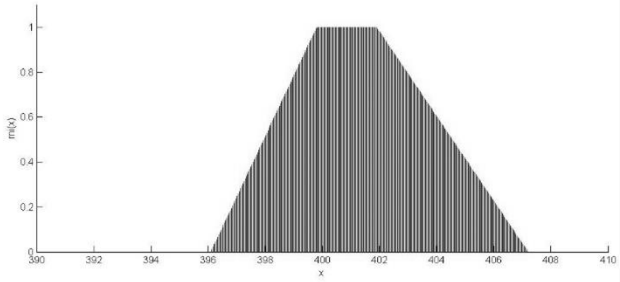


Figure 2 Membership function for portfolio PV.

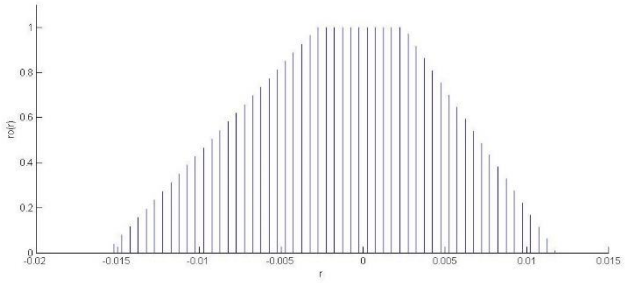


Figure 3 Expected fuzzy return from the portfolio.

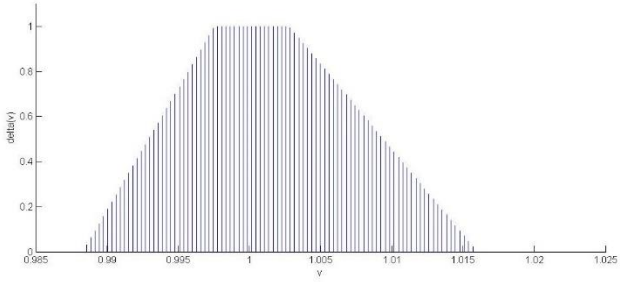


Figure 4 Expected fuzzy discount factor from the portfolio.

Energy and entropy measures, calculated according to (15), (16), (21) and (22) for both portfolio and each asset, are given in Table 2.

asset\measure	A1	A2	A3	A4	π
energy	250	20	325	15	610
entropy	0.33	0.33	0.16	0.14	0.25

Table 2 Simulation results for risk measures of researched assets and their portfolio.

The performed simulation resulted in following dependences:

$$\begin{gathered} d(A_\pi) > d(A_3) > d(A_1) > d(A_2) > d(A_4) \\ e(A_1) = e(A_2) > e(A_\pi) > e(A_3) > e(A_4) \end{gathered} \tag{23}$$

5 Final remarks

Based on performed research we conclude that in case of imprecise present value modelled by discrete fuzzy

number, the construction of the portfolio:

- can increase ambiguity risk,
- can average indistinctness risk.

This means that multiple asset portfolio diversification does not result in minimizing both of the researched risk components, which suggests the sub optimality of using this method and the necessity of creating a new risk minimization and revenue maximization method. Stated conclusions are reflected in analyzed case study and performed simulations. Achieved results suggest purposefulness of presented research and encourage further studies, especially in the area of risk management.

Acknowledgements

The article is financed by the National Science Centre Poland, grant No. 2015/17/N/HS4/00206.

References

[1] Casasnovas J, Vicente Riera J., On the addition of discrete fuzzy numbers, *Proceedings of the 5th WSEAS International Conference on Telecommunications and Informatics*, Istanbul, Turkey, May 27-29, (2006), pp. 432-437

[2] Czogała E. Gottwald S., Pedrycz W., Contribution to application of energy measure of fuzzy sets, *Fuzzy Sets and Systems*, **vol**. 8, (1982), pp.205-214

[3] De Luca A., Termini S., A definition of non-probabilistic entropy in the setting of fuzzy set theory, *Information and control*, **vol**. 20, (1972), pp. 301-312

[4] Dubois D., Prade H., Fuzzy Sets and Systems: Theory and Applications, *Mathematics in Science and Engineering*, **vol**. 144, 1980.

[5] Fang Y., Lai K. K., Wang S., Fuzzy portfolio optimization. Theory and methods, *Lecture Notes in Economics and Mathematical Systems*, **vol**. 609, Springer, Berlin, 2008.

[6] Guixiang Wang, Cheng Lin Wen, A new fuzzy arithmetic for discrete fuzzy numbers, *Fourth International Conference on Fuzzy Systems and Knowledge Discovery*, (2007), pp. 52 - 56

[7] Guixiang Wang, Qing Zhang, Xianjun Cui, 2008, The discrete fuzzy numbers on a fixed set with finite support set, *IEEE Conference on Cybernetics and Intelligent Systems*, (2008), pp. 812-817

[8] Huang X., Risk curve and fuzzy portfolio selection, *Int. J. Production Economics*, **vol**. 106, (2007), pp. 1102-1112

[9] Khalili S., Fuzzy measures and mappings, *J. Math. Anal. Appl*., No. 68. (1979).

[10] Klir G.J., Developments In Uncertainty-based Information, In*: Yovits M. (ed.), Advances in Computers*, **vol**. 36, Academic Press, San Diego, 1993, pp. 255-332

[11] Kuchta D., Fuzzy capital budgeting, *Fuzzy Sets and Systems*, **vol.** 111, (2000), pp. 367-385

[12] Lesage C., Discounted cash-flows analysis. An interactive fuzzy arithmetic approach, *European Journal of Economic and Social Systems*, **vol**. 15(2), (2001), pp. 49-68

[13] Markowitz H.S.M., Portfolio selection, *Journal of Finance*, **vol**. 7, No. 1, (1952), pp. 77-91

[14] Piasecki K., Behavioral Present Value, *SSRN Electronic Journal*, **vol.** 01/2011, (2011), DOI:10.2139/ssrn.1729351.

[15] Piasecki K., Siwek J., Behavioural Present Value defined as a fuzzy number – a new approach, *Folia Oeconomica Stetinensia*, **vol**. 2015, pp. 27-41.

[16] Piasecki K., Siwek J., , Portfel dwuskładnikowy z trójkątnymi rozmytymi wartościami bieżącymi - podejście alternatywne, *Przegląd Statystyczny*, **vol**. 2017 (1).

[17] Siwek J., Portfel dwuskładnikowy – przypadek wartości bieżącej danej trapezoidalną liczbą rozmytą, *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach* (2015), (in review).

[18] Siwek J., Portfel dwuskładnikowy – studium przypadku dla wartości bieżącej danej jako trójkątna liczba rozmyta, *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, **vol**. 141/2015, (2015b), pp. 140-150.

[19] Vicente Riera J., Torrens J., Aggregation functions on the set of discrete fuzzy numbers defined from a pair of discrete aggregations, *Fuzzy Sets and Systems*, **vol**. 241, (2014), pp. 76-93

[20] Vicente Riera J., Torrens J., Using discrete fuzzy numbers in the aggregation of incomplete qualitative information, *Fuzzy Sets and Systems*, **vol**. 264, (2015), pp. 121-137

[21] Voxman W., Canonical representation of discrete fuzzy numbers, *Fuzzy Sets and Systems*, **vol**. 118, (2001), pp. 457-467

[22] Ward T.L., Discounted fuzzy cash flow analysis, *1985 Fall Industrial Engineering Conference Proceedings*, Berkeley, (1985), pp. 476-481

[23] Zadeh L.A., Fuzzy sets, *Information and control*, **vol.** 8, (1965), pp. 338-353

Estimates of Regional Flows of Manufacturing Products in the Czech Republic

Jaroslav Sixta[1], Jakub Fischer[2]

Abstract. The Czech economy is crucially dependent on manufacturing industry since it is the most important sector of the economy. In 2015, the share of manufacturing on value added is about 27% and exports of manufacturing products takes 76% of GDP. Producers of manufacturing products place their factories disproportionally in the Czech Republic. Therefore the impact on regional economy is different. However, the production of manufacturing products requires lots of different inputs and therefore the links between regions are tighter that can be directly observable. Moreover, the inputs for manufacturing industry contain also services and the distribution of created value added is more complex. The aim of the paper is to analyse the links in Czech manufacturing industries using regional input-output tables for fourteen regions of the Czech Republic. These tables were compiled for 2013. By the Newton gravity method, they were arranged into the inter-regional model as a matrix with 1148 rows and columns. The paper brings the estimates of both the dependency of regional value added and regional employment on manufacturing industry.

Keywords: Input-Output, Regional, Models.

JEL Classification: C67, O11
AMS Classification: 65C2

1 Introduction

The Czech economy is known for its high share of manufacturing among developed countries. This results from the long tradition of mining, manufacturing and energy production on the Czech territory. In modern word, the statistical description of production is not straightforward. Times, when simple economic indicators such as industrial production index, quarterly gross value added etc. has passed. The amount of the production measured either by the system of accounts (SNA) or short term statistics (STS) does not provide adequate information. Due to the complexity of economic relationships and processes, production or better value added should be carefully split between the part originated in the domestic territory and the part originated abroad. On the national level, this information is provided by the OECD within the TIVA project (OECD, 2015) and on the regional level within sophisticated regional input-output tables.

Even though the export of the Czech economy takes a high share of gross domestic product, it must be mentioned that significant part of export is formed by imported products. These intermediates contain parts, finished products, services but also natural resources such as crude oil or gas. Fundamental question is therefore the dependency of exports on imports and the value added embodied in exports.

Similarly to the national level, on the regional level manufacturing regions are interlinked with other regions, usually big cities such as Prague and Brno. Company headquarters seated in these cities are providing services to their local subsidiaries. Official statistics provides regional accounts but regional input-output tables must be prepared by academic research in the Czech Republic, see Kahoun and Sixta (2013). Despite we focuse on the description of the production side of the economy but the use side can be described similarly, see Kramulova and Musil (2013).

The aim of the paper is to analyse and trace the flow of manufacturing product both on national and regional level. It is shown that similar process presented on the national level can be prepared on the regional level by using regional input-output tables. With respect to the dependency of the Czech Republic, we focused also on the production of transport equipment.

[1] University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, e-mail: sixta@vse.cz.
[2] University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, e-mail: fischerj@vse.cz.

2 Data and Methodology

Regional product flows can be considered on the national level or at the sub-national level. The first concept refers to the globalised world where the economic borders between countries are not very relevant. The second concept allows decomposition of countries´ figures into regions, regional input-output tables, see Miller and Blair (2009).

Foreign participation of the country on the international market is usually measured by the share of exports on gross domestic product. From the theoretical point of view, export should be directly linked to output. Even that seems logical, lots of obstacles resulting in using GDP as a suitable denominator. It is difficult to ensure international comparability of output since it is dependent on frequently changing number of units, level of aggregation and statistical procedures. Moreover, implementation of ESA 2010 brought the measurement of re-exports and exports can occur even if the domestic output is missing. The data comes from Eurostat database.

Statistical description of the flows of products between countries is used by the OECD in TiVA database (Trade in Value Added). Among lots of data, this database includes countries' IOT that allows estimates of the trade in value added. For n = 62 countries (economies) and k = 34 industries TiVA contains Inter-Country Input-Output (ICIO) system. This information provides important knowledge about the participation of a country in foreign trade and the benefits resulting from it. We use two important indicators, *domestic value added embodied in gross exports* of a country c, partner p and industry i ($D_{c,p,i}$) and *foreign value added content of gross exports* of a country c (F_c):

$$D_{c,p,i} = V_c \cdot B_{c,c} \cdot E_{c,p,i} \tag{1}$$

$$F_c = V \cdot B_{(c)c} \cdot E_{c,i} \tag{1}$$

where

V_c	1 x k row vector with domestic value added shares of output for each industry i,
V	1x (n x k) value added to output ratio by industry and country,
$B_{c,c}$	k x k diagonal block matrix of B (Global Leontief inverse) representing total domestic gross output required for one unit increase of country c's demand,
$B_{(c),c}$	column block of B corresponding to country c, with the row block corresponding to c being zero,
$E_{c,p,i}$	k x 1 vector of gross exports from country c to country p for industry i,
$E_{c,i}$	k x1 vector of total exports of country c for industry i.

The framework of the mathematical background is given by the Leontief's equations. Matrix algebra of the input-output framework provides a wide range of potential use. The most important is the equality of sources and uses and therefore the crucial aim of both classic and extended input-output analysis is finding for a new equilibrium.

On the national level, the analysis is similar. It is not easy to identify the share of the region in the total value added of manufacturing products. For these purposes, we used regional input-output tables presented within the project conducted by the Department of Economic Statistics "Regionalization of Estimated Gross Domestic Product by Expenditure Method", no. 13-15771S, see Sixta and Vltavská (2016). These regional input-output tables describe the Czech economy on the level of 14 regions (NUTS 3 level). These tables were prepared as symmetric input-output tables for 82 products at basic prices. The tables are based on similar procedures as in Többen and Konenberg (2015). The set of individual tables was arranged into inter-regional model, large matrix with 14 x 82 rows and columns. Inter-regional model was prepared by Newton gravity method by Šafr (2016). The methodology is deeply described in Šafr and Vltavská (2017). Since input-output tables are compiled every 5 years (years ending by 0 and 5), the newest input-output tables will be compiled for 2015 in June 2017. The Czech Statistical Office also sometimes compiles IOT for the period between, in this point for 2013. Therefore the latest available data source for RIOT is 2013.

3 Results on the National Level

The Czech Republic is very opened economy and its exports takes 76.9% of GDP in 2013, see Table 1.

		1995	2000	2005	2010	2013
Exports	**Czech Republic**	40.6	48.3	62.3	66.2	76.9
	Hungary	39.3	66.8	62.8	82.2	86.0
	Poland	23.0	27.2	34.6	40.1	46.3
	Slovakia	56.7	54.1	72.0	76.3	93.8
Imports	**Czech Republic**	43.7	50.2	60.0	63.1	71.1
	Hungary	39.2	70.5	65.1	76.9	79.0
	Poland	20.7	33.6	35.7	42.1	44.4
	Slovakia	54.6	56.6	76.6	77.8	89.6

Table 1 Share of exports and imports on GDP (%)

Manufacturing products take 80% of Czech exports. Even such high shares express importance of export for the Czech economy, the situation is more complex. In globalized world, exports generate imports in many cases. This is exactly valid for the Czech Republic and for Slovakia, Hungary and Poland, as well.

The most important part of exports of these countries is composed from manufacturing products. It reflects the tradition of manufacturing industry in the Central Europe and foreign investments in the last decade. The most significant increase of manufacturing products' exports is in the Czech Republic and Poland, see Figure 1. With respect to the size of the economy, the dependency of the Czech economy on foreign demand is clearly visible. But for estimates of the value added embodied in export, input-output approach is needed.

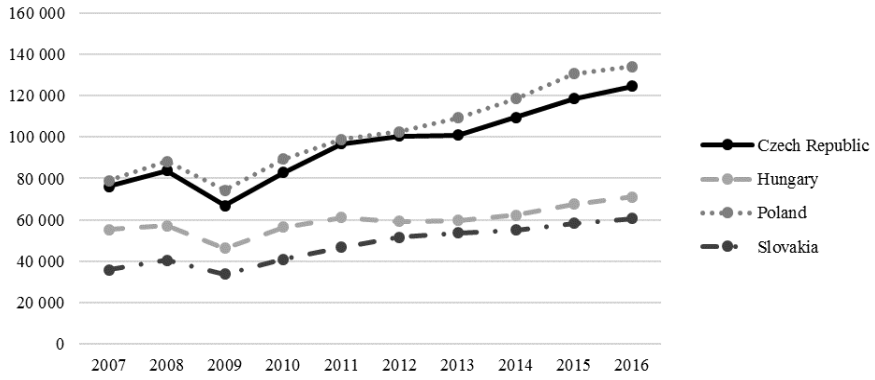


Figure 1 Exports of manufacturing products, mil. EUR

The input-output approach presented by the OECD to the estimates of the components of exports belongs to the most valuable statistical information. Since TiVa contains the latest figures for 2011, we can compare exports with domestic value added for the period 1995–2011 only. The ability of a country to transform foreign demand (exports) into domestic gross value added is expressed by the embodied gross value added in exports. It means that the exports are composed from two parts, domestic value added and imported value added (value added created outside of domestic territory). In all mentioned countries, the share of domestic value added has been decreasing, see Figure 2. The highest share of domestic value added is observed in Poland, where the indicator dramatically declined from 75.9% to 52.8% between 1995 and 2011. On the contrary, in the Czech Republic the indicator was relatively stable since it decreased from 48.8% to 46.9%. In Hungary and Slovakia, the share of domestic value added is significantly lower and even though it decreased from 43.2% (44.3% respectively) to 39.4% (39.6% respectively).

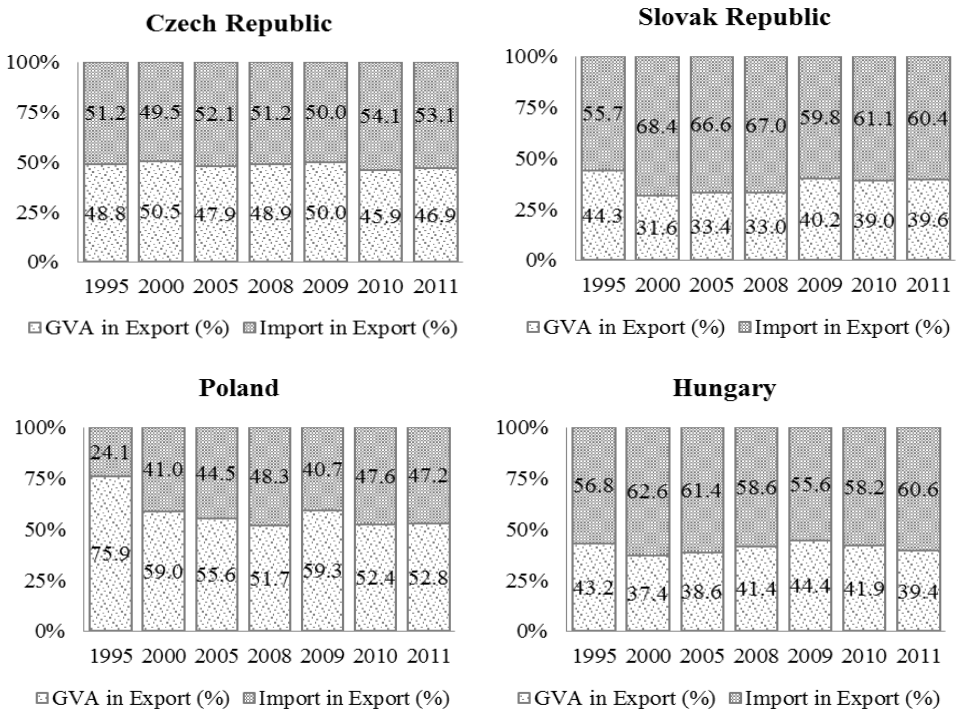


Figure 2 Domestic and foreign components of exports

4 Results on the regional level

Similarly to the national level, the contribution of a particular region to gross value added can be estimated by using input-output approach. Regional input-output tables allow to construct the flows of products and the contribution of the regions to overall gross value added. In 2013, the share of gross value added of manufacturing products takes 24% and amounts of 32 bn. Eur. The export of manufacturing products amounts of 95 bn.

In general, modern regional accounts compiled in line with ESA 2010 are published by the Czech Statistical Office. The following Table 2 illustrates the composition of gross value added by the regions in the Czech Republic for 1995–2013.

Region	Label	1995	2000	2005	2010	2013
Hlavní město Praha	PHA	20.0	22.7	24.5	25.7	24.7
Středočeský kraj	STC	9.9	10.9	10.4	10.6	11.0
Jihočeský kraj	JHC	5.8	5.7	5.5	5.1	5.2
Plzeňský kraj	PLZ	5.2	5.0	5.1	5.0	5.1
Karlovarský kraj	KAR	2.8	2.5	2.3	2.1	2.0
Ústecký kraj	UST	7.7	6.6	6.6	6.3	6.1
Liberecký kraj	LIB	3.8	3.8	3.5	3.2	3.2
Královéhradecký kraj	KRH	4.9	5.0	4.6	4.6	4.5
Pardubický kraj	PAR	4.4	4.2	4.0	4.0	3.9
Kraj Vysočina	VYS	4.2	4.1	4.1	3.9	4.1
Jihomoravský kraj	JHM	10.3	10.0	9.9	10.3	11.1
Olomoucký kraj	OLM	5.2	5.0	4.7	4.6	4.7
Zlínský kraj	ZLN	4.9	4.8	4.6	4.7	4.7
Moravskoslezský kraj	MRS	10.9	9.6	10.3	9.8	9.7

Table 2 Composition of gross value added by the regions of the Czech Republic, %

The contribution to the output of manufacturing products is presented on Figure 3. These estimates come from regional input-output tables for 2013. From this perspective, the most important regions are Středočeský and Moravskoslezský kraj, they both are closely connected with transport equipment industry, mainly automo-

tive[3]. But the situation is more complex since inputs for automotive industry are provided by other regions, as well.

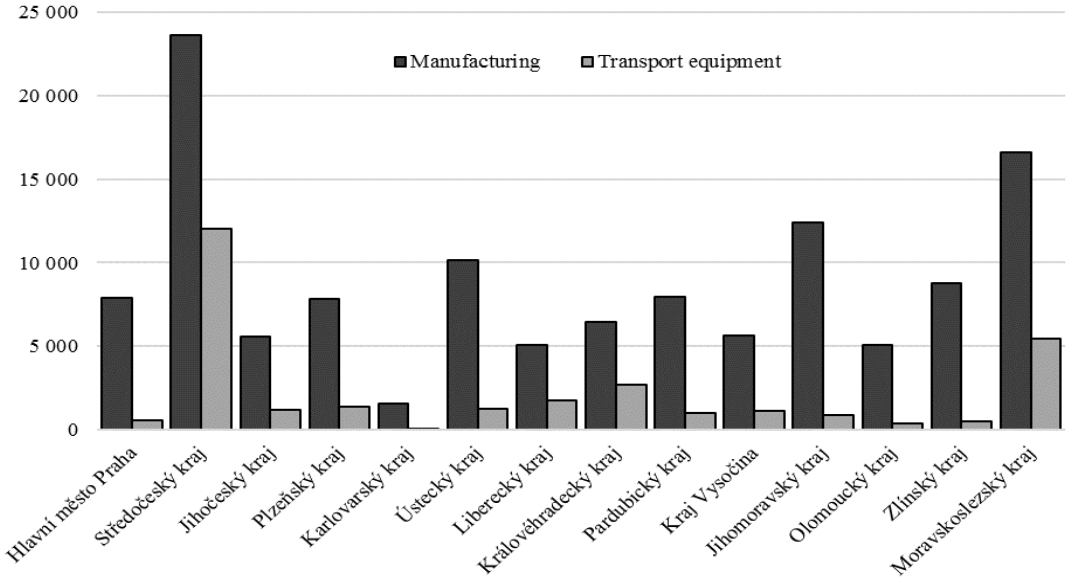


Figure 3 Output of manufacturing products by regions, mil. EUR

On the following flowchart (Figure 4), output flows of manufacturing industries are presented between all 14 regions of the Czech Republic. These flows reflect the connection between the regions in manufacturing products. We can identify strong connections (bold lines) between Jihočeský (JHC), Středočeský (STC), Zlínský (ZLN) and Moravskoslezský kraj (MRS). On the contrary, the capital city Prague (PHA) is nearly proportionally connected with all other regions.

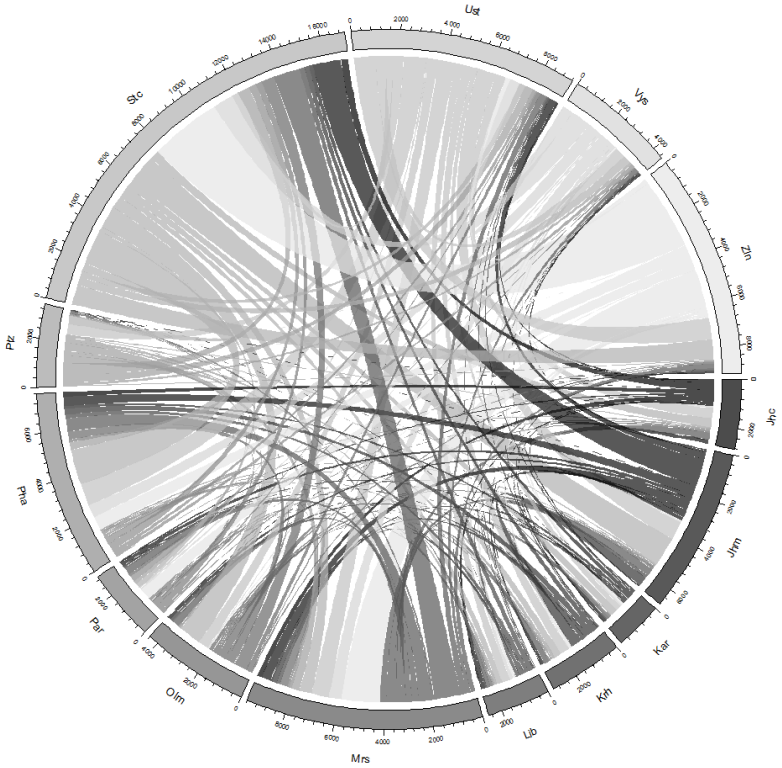


Figure 4 Inter-regional flows of manufacturing products, Czech Republic, 2013.

Note: Region labels are explained in Table 2.

[3] Another suitable indicator could be based on physical units, e.g. Zbranek and Fischer (2016) but it is not broken down by the regions up to now.

5 Conclusion

Manufacturing industry is the most important factor of economic growth in the Czech Republic. This means that not solely production of manufacturing products provide income for the Czech residents but also many delivering and connected industries. The question, how important is the output of manufacturing industry for the Czech Republic is not easy. Such issue contains at least two approaches.

The first is aimed at country as an element of economic blocks as performed by the OECD. It leads to the estimates of value added embodied in exports. For the Czech Republic, the share of value added in exports is lower than 50% for the whole available period (1995-2011). That reflects the situation of globalized word and intensive foreign investments in the Central Europe.

The second approach relates to a sub-national level when country totals are distributed among regions. For these purposes, regional input-output tables are used. They provide valuable information about regional output and intermediates broken down by products. The amount of information hidden in regional input-output tables is multiplied when these individual tables are arranged into inter-regional model. It means that the links between regions can be described by product flows. From the analysis of inter-regional flows, the dependency between Jihočeský, Středočeský and Moravskoslezský kraj is seen.

The analytical potential of advanced input-output models is very high. Such information should be used as basis for political decisions, e.g. SWOT analysis. It may provide valuable information in terms of the impact of foreign investments on the level of employment, wages and regional gross domestic product.

Acknowledgements

Supported by the grant No. VŠE IGS F4/28/2017 and Institutional Support for Long Period and Conceptual Development of Research and Science at Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] EUROSTAT: *European System of Accounts (ESA 2010)*. Eurostat, Luxembourg, 2013.
[2] Kahoun, J., Sixta, J.: Regional GDP Compilation: Production, Income and Expenditure Approach. *Statistika*, **4** (2013), 24–36.
[3] Kramulová, J., Musil, P.: Experimentální odhad složek výdajové metody regionálního HDP v ČR. *Politická ekonomie* (2013), 814–833.
[4] Miller, R.E., Blair, P. D.: *Input-Output Analysis: Foundations and Extensions*. Cambridge University Press, 2009.
[5] OECD, *TiVA 2015 indicators – definitions*, October 2015, Paris.
[6] Sixta, J., Fischer, J.: Possibilities of Regional Input-Output Analysis of Czech Economy. In: *Mathematical Methods in Economics* [CD-ROM]. Liberec, 06.09.2016 – 09.09.2016. 762–767. ISBN 978-80-7494-296-9.
[7] Sixta, J., Vltavská, K.: Regional Input-output Tables: Practical Aspects of its Compilation for the Regions of the Czech Republic. *Ekonomický časopis* **64(1)** (2016), 56–69. ISSN 0013-3035.
[8] Šafr, K. Allocation of commodity flows in regional Input-Output tables for the Czech Republic. *AMSE* (2016), 328–340.
[9] Šafr, K., Vltavská, K.: Illustration of Single-Regional and Inter-Regional Approach in Regional Input-Output Analysis. *Statistika* **97(1)** (2017), 18–31. ISSN 0322-788X
[10] Többen J., Konenberg. T.H.: Construction of Multi-Regional Input-Output Tables Using the Charm Method. *Economic Systems Research* **4** (2015), 487-507.
[11] Zbranek, J., Fischer, J.: Konstrukce a využití časových input-output tabulek pro hodnocení produktivity práce v podmínkách České republiky. *Politická ekonomie* **62(6)** (2014), 769–784. ISSN 0032-3233

Software for Changepoints Detection

Hana Skalska[1]

Abstract. The paper concerns with freely available software as the possibility to solve the task of detection and statistical evaluation of changes in occurrences (counts, rates, or proportions) of some event, when data is collect over time or in dependence on quantitative attribute. Off-line data analysis is assumed and common principles of change point problems solution are described. Freely available software implementing piecewise linear regression, joinpoint model, segmented regression, changepoints, and adaptive regression splines is selected. The article describes and compares the main features of selected software. Public available data (unemployment rates) are used to compare solutions implemented in selected R packages (earth, changepoint, and segmented) and in the stand-alone Joinpoint Regression Program (JP). The results are not identical partly due to the properties of data and mostly due to different approaches to CP solutions. JP support user to avoid some problems with data collected in time, with difficulties related to assumptions of parametric linear model, or multiple hypotheses testing.

Keywords: Breakpoints, joinpoints, JP software, R packages, earth, segmented, changepoint.

JEL Classification: C50
AMS Classification: 62J05

1 Introduction

The problem arises from the practice and the necessity to recommend software eligible to detect changes in a series of observations (counts, occurrences, rates, proportions, or other measurable characteristics) collected in time or in dependence on any quantitative attribute. Detection and statistical evaluation of changes is frequently demand and solved task of practical importance. In economics, finance, demographics, epidemiology, and other fields, the task is to detect changes in model (linear, broken-line, or piecewise linear) or in parameters of the process. Unknown change points (breakpoints) in trend or mean response of observed characteristics Y on time or other quantitative predictor Z have to be estimated from a set of observations (z_1,y_1), (z_2,y_2), …(z_n,y_n), $z_1 \leq z_2 \leq \ldots \leq z_n$.

Change point (CP) is the point in which an ordered sequence of data $(y_1, y_2, \ldots, y_n)$ changes its statistical properties. A changepoint occurs when there exists a point $\tau \in \{1, 2, \ldots, n-1\}$ such, that the statistical properties of $\{y_1, y_2, \ldots, y_\tau\}$ and $\{y_{\tau+1}, y_{\tau+2}, \ldots, y_n\}$ are different in some characteristics [13]. Regression change points connect subsequent segments (usually adjacent linear pieces) that differ each other in levels of parameters of observed quantitative attribute. Anomaly (outlier) detection solve similar problem detection of data points that do not conform to expected behaviour. Methods of change points and anomaly detection depend on nature of data, application domain, and knowledge about distribution and parameters of the process [9].

The objective of the article is mapping and brief overview of change point detection problems and statistical approaches to their solutions. Attention is stressed to piecewise linear regression (segmented, broken-line), joinpoint model, and adaptive regression splines model. The contribution of the article is in effort to categorise CP tasks from an applied point of view and present possibilities of selected freely available software for change point detections. An example of public available data specifies and presents main features of selected software.

2 Common principles of change point detection problems

Ross [25] distinguishes two main categories of change detection problems. *Batch detection* problem (Phase I) and *sequential detection* problem (Phase II). *Batch problem* (or off-line) assumes fixed length of n retrospective observations. Decision about occurrence of a change point is based on all observations (z_1,y_1), (z_2,y_2), …(z_n,y_n), realizations of i.d. random variables $Y_1, Y_2, \ldots, Y_n$. *Sequential problem* is processed on sequential data $Y_1, Y_2, \ldots$, collected in time series or signal streams. The next (new) observation received in a sequence is qualified as consistent with the previous sequence or inconsistent, potential to belong to the next segment with changing parameters. Decision is based on an appropriate test statistic, penalty, and decision criteria. Phase II problems are typical for big data or signal streams, where interest in multiple CP detection is growing with increasing availability of

[1] University of Hradec Kralove, Faculty of Informatics and Management, Department of Quantitative Methods and Informatics, Rokitanskeho 62, 500 03 Hradec Kralove, hana.skalska@uhk.cz.

sensor data (genomic data, bioinformatics, signal processing, network monitoring, intrusion detection, fraud detection, etc.). Some algorithms are able to solve both tasks.

Different statistical and non-statistical approaches exist to solution of retrospective CP problem. The paper concentrates to frequentist solutions. Parametric methods assume that observed data follow an appropriate distribution. Change point (joinpoint, piecewise) regression solution estimates linear segments. Maximum likelihood estimates (MLE) of parameters of the model allow specify threshold for a test criterion, determine a change and emphasise the continuity of the linear segments [15, 16].

Increasing number of R [24] packages offer various methods for change detection. The main categories corresponding to our interest are broken-line regression that is available in the R package **segmented** [22, 23], multivariate adaptive regression splines MARS [4, 26] implemented in the R package **earth** [19, 20], and algorithms for multiple changepoints detection available in the R package **changepoint** [14].

Piecewise linear regression and Join Point Regressin Programme (JP)

Regression model assumes that a function with unknown regression dependence describes the process. The function is approximated by linear pieces (segments) connected at changepoints (joinpoits, break-points, transition points). Piecewise linear models are preferred in applications as they are simple, easily interpreted, and available software offers these methods. Piecewise regression describes trend as a sequence of straight line segments connected at joinpoints with significant changes in parameters The joinpoint regression model can be written as

$$E[y \mid z] = \beta_0 + \beta_1 z + \delta_1 (z - \tau_1)^+ + \ldots + \delta_M (z - \tau_M)^+$$

where M is number of break points, and τ_i are breakpoints, i = 1, 2, …, M, $a^+ = \max(0, a)$.

Different articles solved the problems introduced with Julious [12], when a CP location is unknown. Problems are in location estimate, significance in change, and parameter estimates. Estimates of two unconstrained lines at a point have to be adjusted to satisfy model continuity of linear segments [15, 16] and to form continuous function [17]. Statistic F tests the hypothesis about significant change in slope and compares residual sum of squares (RSS) of the two-lines and one-line models. Exact F distribution for this statistic is unknown if changepoint location is unknown. Simulations and bootstrap methods can estimate decision threshold [12].

The MLE of ratio statistics, permutation tests, and other principles are available to estimate a change-point location. Typical methods and algorithms test the H_0 (no change point) against the H_1 (there are M change points). When H_0 is rejected, the test of H_0 against M – 1 points continues and the process repeats. Sequential tests control the over-fitting. Permutation procedure tests all possible locations of predetermined discrete CP locations in data and searches optimum of a cost function over all possible combinations [14]. Computational efficiency, multiple testing, and statistical properties of detected change points are difficulties related to this process. Different solutions overcome these problems. An overview of approaches to change point detection, statistical tests for changes in location model, critical values obtained through permutation tests, and an exact formulation and solution of some theoretical principles of permutation tests are in [1,7]. The advantage of permutation tests to CP solutions with application in econometrics describes [3]. Kim et al. [15] derived conditions and specified methods under which permutation procedure is consistent in estimate of the number of change points. Simulation study concluded [15] that BIC criterion is more sensitive to picking up smaller changes while the permutation procedure is preferred when the goal of an analysis is a parsimonious model.

Change point (joinpoint) regression solution is available in **Join Point Regressin Programme** (**JP**) [11]. **JP** implements methods [8, 14, 15, 16] and allow to select permutation tests, BIC criterion, or modified BIC. A comparison of **JP** and Bayessian approaches with application to cancer mortality data is presented in [5].

Multivariate regression with adaptive splines

Multivariate adaptive regression splines (MARS, trademark of Salford System, calls mars:mda in the **earth** package [4, 20]) built multivariate non-linear model upon the recursive partitioning. Method is able to solve high dimensional problems. MARS estimates piecewise linear function. In one-dimensional problem, the function at every observed z becomes the form

$$f(x) = \beta_0 + \beta_1 (z - \tau_1)^* + \beta_2 (z - \tau_2)^* + \ldots + \beta_M (z - \tau_M)^*$$

where M is number of change points (knots), τ_i are the best knots (nodes), that needs to be determined from the data (adaptive model). The component $(z - \tau)^*$ is either negatively $(z - \tau)^+$, or positively $(z - \tau)^-$ truncated function. For any number a, $a^+ = \max(0, a)$, and $a^- = (a^+ - a)$. In the forward steps, MARS selects the best predictors, the best knots, and finds set of candidate models. The backward step uses the Generalized Cross Validation (GCV) to select the best final model. GCV reflects the model complexity and the lack of fit [26]. User specifies the maximum

number of segments. This specification depends on data dimension but the impact of selected number is relatively small in the model selection [26].

Changepoint solution

Description of the R package **changepoint** and an overview of packages convenient to various (mostly) sequential detection problems can be found in [13]. This package makes available a broad range of single and multiple CP solutions. Multiple changepoint problem assumes that data are arranged in ordered sequence ($y_1, y_2, \ldots, y_n$) and there exists a change point in time $\tau \in \{1, 2, \ldots, n-1\}$ and extends this idea to M changepoints at M ascending positions $\{\tau_1, \tau_2, \ldots, \tau_M\}$. Then M changepoints split data to M + 1 segments, each segment is summarized by a set of parameters. The changepoint package implements three algorithms (binary segmentation, segment neighborhoods, and PELT). Pruned exact linear time (PELT) algorithm was proposed recently [13].

Segmented solution

Muggeo [21] developed different approach to regression with unknown position of change points. Algorithm in the R package **segmented** [22, 23] needs the initial guess ψ_{est} for a break point. Problem formulation assumes segmented relationship between the mean response $\mu = E[Y]$ and predictor variable Z observed for i = 1, 2, …, n. Segmented model is built by adding $\beta_1 z_i + \beta_2 (z_i - \psi)_+$ term in the linear predictor, $(z_i - \psi)_+ = (z_i - \psi) \cdot I(z_i > \psi)$, where I(.) is an indicator function. Model fits the linear term $\beta_1 z_i + \beta_2 (z_i - \psi_{est})_+ + \gamma\, I(z_i > \psi_{est})^-$ iteratively, and $I(\cdot)^- = -I(\cdot)$. Parameter γ is understood as re-parametrisation of ψ and accounts for the breakpoint estimate. The breakpoint value is updated iteratively.

Other approaches

Change detection problem is extended to multivariate methods too. Many algorithms are available to detect single or multiple CPs for big data (or data streams). One class of approaches minimises costs over segmentation and uses dynamic modelling principles [2], some methods speed up algorithms by segmenting data and implement pruning of solution space [13,18]. Another approach suggests sparsity and L_1 regularization techniques and constructs the CP model on the base of all possible alternatives admitted by data [17].

3 Software specifications

The statistical software package **Joinpoint Regression Program (JP)** version 4.4.0.0 [11] fits data into the joinpoint models. Sequence of permutation tests (Monte Carlo method) identifies CPs in trend. The procedure is based on a sequential application of likelihood ratio type tests [14]. Number of min and max of JPs are specified by user. Simple linear or log-linear model, modelling criteria, final model selection methods (permutation test, BIC, or Modified BIC), and parameters of permutation test of significance are specified by user. Adjustment procedure implemented since version 3.0. uses method less conservative than Bonferroni correction used in older versions. This version of JP allows pair-wise comparisons (tests of parallelism or coincidence) between two data sets. Standard errors of dependent variable can either be provided to JP or can be calculated by JP (when dependent variable is count and follow Poisson distribution, or when necessary data are given to the JP). Homoscedastic or heteroscedastic (WLS) regression fits model in dependence on error patterns [10]. Grid search and models with autocorrelation errors are possible.

The R package **earth** builds lm and Glm standard models (mars), uses pruning and cross-validation. The package has a capability to solve big data problems [20]. The Cross validation (CV) principle allow to estimate standard error of the model and confidence limits, helps to select the best model and to control overfitting. The Generalized RSquare (GRSq) measure the ability of generalization the model. General Cross Validation (GCV) is penalty per knot, the value −1 (no penalty) is allowed. GRSq normalises GCV and supports selection the number of effective parameters in the model. Variance models in **earth** allow estimate of prediction intervals under assumption that errors are independent and possibly heteroscedastic [19].

The R package **changepoint** [13] detects multiple changes within a time series or sequence. Package implements several existing multiple changepoint search algorithms and introduces PELT (pruned exact linear time) that provides an exact segmentation. Algorithm is focused to changes in mean, to changes in variances, and to changes in mean and variance. In principles, PELT is similar to the neighbourhood binary segmentation algorithm. Test statistic assumes distribution of data (possible choices are normal, exponential, Gamma, or Poisson).

The R package **segmented** [22, 23] solves univariate and multivariate lm and GLM models with broken-line relationships. Estimate of the model uses grid-search algorithm and translates CP detection problem into the standard linear framework, uses re-parametrisation, and estimates linear model iteratively by updating initial estimate. Bayessian Information Criterion (BIC) helps to estimate the number of breakpoints, and the Davies test estimates

location a breakpoint and checks its significance. Dependence of resulting solution on the guess of initial solution was mentioned in [6].

3.1 Data and results

The data collected monthly describes the ratio (in %) of job applicants aged 15 to 64 in the population of the same age at the Czech Republic (http://portal.mpsv.cz). The unified method of collection and calculation is used since 2004. Observed data (points on *Figure 1*) correspond to time period from 01_2005 to 12_2016. Autocorrelation, influence of seasons, and possible changes in trend can describe the series. The analysis is focused to detection changes in trend.

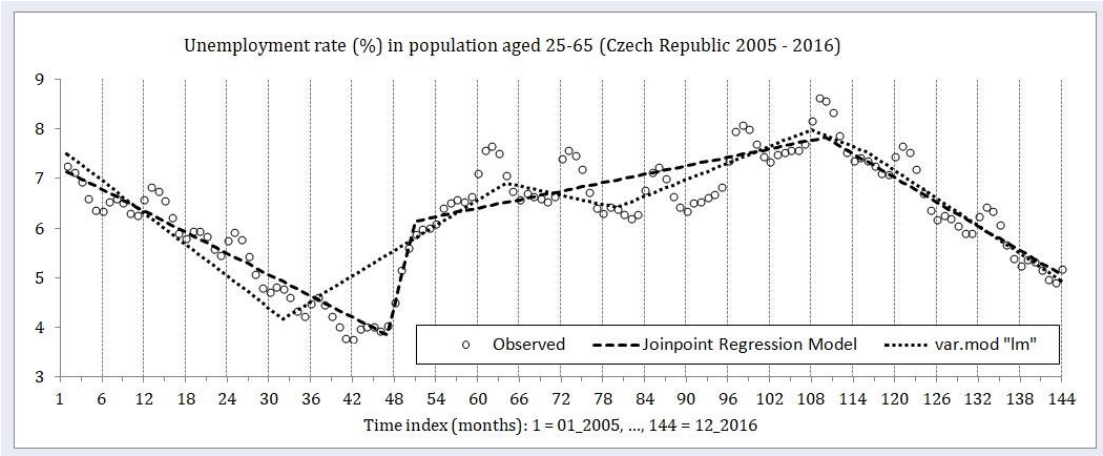


Figure 1 Comparison of **JP** and **earth** piecewise regression models

Figure 1 combines results of two models (coefficients in *Tables 1* and *2*). Both software produce graphs but *Figure1* allow visual comparison of differences in two models. **JP** detects three JPs with indexes 47, 51, and 110 (months 11_2008, 03_2009, and 02_2014). Corresponding 95% confidence intervals of CPs locations are (40 – 48), (49 – 63), and (105 – 123). Parameters and standard errors of the **JP** piecewise model are in *Table 1*. The options used for **JP** software settings: Homoscedastic sample errors, grid search, fitting the autocorrelated errors based on data, MBIC criterion for selection of the best model. Time creating the model was 78 min. Final model error SSE = 23.38. Slopes as well as differences between slopes of adjacent lines are different from zero at significance level $\alpha = 0{,}05$. Log-linear model (available in **JP)** resulted with identical CPs (not presented here).

	Intercept **1**	Intercept **2**	Intercept **3**	Intercept **4**	Slope **1**	Slope **2**	Slope **3**	Slope **4**
Parameter Estimate	7.210	-23.193	4.685	16.803	0.072	0.575	0.029	-0.082
Standard Error	0.239	6.348	0.409	1.376	0.008	0.130	0.005	0.011
Index range	1 – 47	47 – 51	51 – 110	110 – 144	1 – 47	47 – 51	51 – 110	110 – 144

Table 1. Parameters and standard errors of **JP** model.

Object model (varmod.method = "lm") from variance model in the R package **earth** estimates prediction intervals and the analysis of errors (*Figure 2*). Parameters of the general model are in *Table* 2. Confidence limits and visualised residuals are in *Figure* 2 (with three detected outlier data points 45, 46, and 47). Model summarises SSE = 35.42 (bigger to the JP model), values GRSq = 0.76, RSq = 0.80, and CVRSq = 0.77 indicate satisfactory quality and stability.

(Intercept)	h(x-32)	h(x-64)	h(x-80)	h(x-108)	h(116-x)	h(x-116)
-4.82704	0.193	-0.117	0.087	-0.113	0.107	-0.143
standard error	0.013	0.018	0.020	0.036	0.008	0.038

Table 2. Parameters and standard errors of mars model from the R package **earth**

Results from the R package **changepoint** v.2.2.2. are presented in *Figure* 3. Method cpt.mean resulted in three CPs locations (27, 50, and 124) with mean values 6.27, 4.43, 7.06, and 5.76 on segments. Method cpt.meanvar is consistent in locations of means with cpt.mean, and adds segments with different variances with CPs at 16, 27, 49, 54, 95, 124, and 135. Mean values on segments are 6.62, 5.76, 4.38, 5.92, 6.78, 7.67, 6.18, and 5.25, and variances 0.086, 0.032, 0.160, 0.029, 0.163, 0.185, 0.030, and 0.049. Multiple segments (horizontal lines) on *Figure* 3 show CPs location, size of segment and location of mean value.

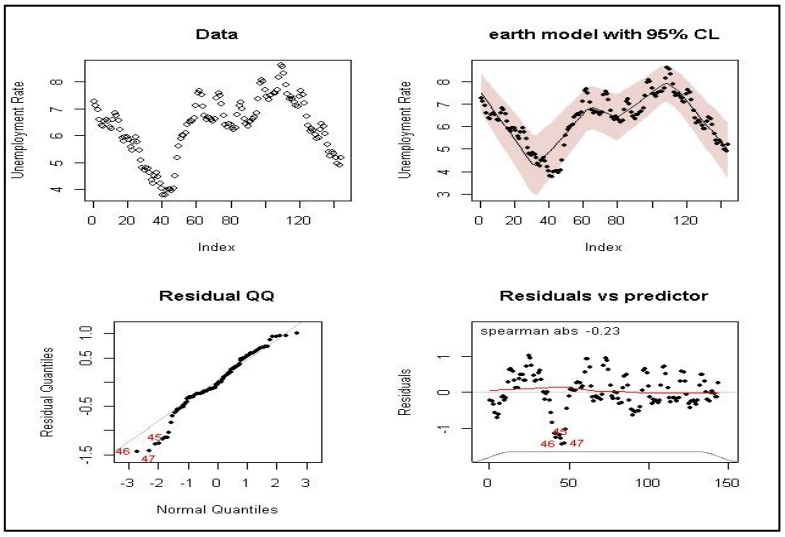


Figure 2 Model **earth** and residuals
Top left to right: Data; Model **earth** with 95% CL; *Bottom left to right*: QQ plot; Residuals and outliers

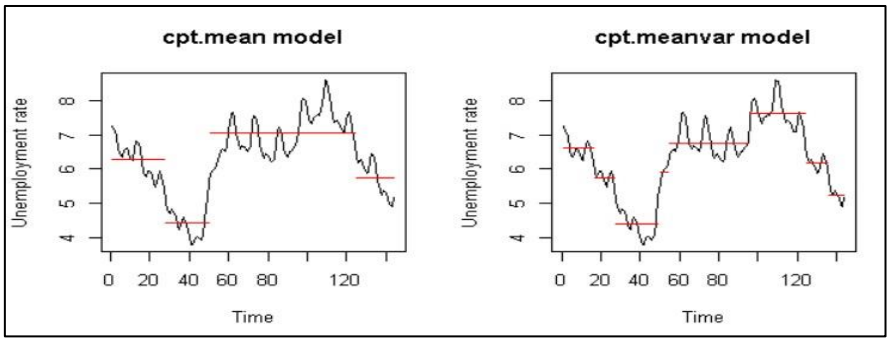


Figure 3 Segments detected with R package **changepoint**

Finally, the results of the R package **segmented** are commented briefly. Function *segmented* in this package needs starting guess for CPs and iteratively finds the best segments. The CP values from **earth** solution were used as the starting guess for function *segmented*. CPs was detected on four locations (43.7, 60.5, 81.9, and 109.9), close to the earth solution with exception of the first point, which is close to the JP solution. Consecutive test of significance with Davies test recommends only the first CP as having significant differences in segment slopes.

4 Conclusion

The presented results are not identical. The JP software and R packages get different results conditioned partly due to the properties of data and mostly due to different approaches to CP solutions, and to different possibilities of treating some common problems with data collected in time. The JoinPoint Regression Programme systematically treats many difficulties following from the nonconformity of data with assumptions of parametric simple linear regression models, and from multiple hypotheses testing. The JP software is designed for an analysis of pre-processed data with size in tens of units, implemented methods are more supported with statistical theory. In opposite, **earth** and **changepoint** packages are designed for big data, and their principles of building and selection the model

combine statistics with computer science methods. Visual and descriptive components of results are available with all presented software. The most important aspects in selection will be the size of data, and the aim of the task (derived from the problem that is solved).

Acknowledgements

This article was partially supported by Faculty of Informatics and Management, University of Hradec Kralove.

References

[1] Antoch, J. and Hušková, M.: Permutation tests in change point analysis. *Statistics & Probability Letters* **53** (2001), 37-46.
[2] Antoch, J. and Jarušková, D.: Testing for multiple change points. *Comput Stat* **28** (2013), 2161 – 2183. doi:10.1007/s00180-013-0401-1.
[3] Černý, M.: K odhadu volatility finančních řad při oceňování derivátů. *Acta Oeconomica Pragensia* **16** (2008), 12-21.
[4] Friedman, J. H.: Multivariate Adaptive Regression Splines. *Ann. Statist*. **19** (1991), 1-67.
[5] Ghosh, P., Huang, L. and Tiwari, R. C.: Semiparametric Bayesian approaches to joinpoint regression for population-based cancer survival data. *Computational Statistics and Data Analysis* **53** (2009), 4073-4082.
[6] Hudecová, Š.: Jak na odhad joinpoint regrese. *Inf Bull České Stat Společnosti* **22** (2011), 7–20.
[7] Hušková, M.: Permutation principle and bootstrap in change point analysis. *Fields Institute Communications* **44** (2004), 273-291.
[8] Hyune, H. J., Binbing, Y. and Feuer, E. J.: Selecting the number of change-points in segmented line regression. *Statistica Sinica* **19** (2009), 597-609.
[9] Chandola, V., Banerjee, A., and Kumar, V.: Anomaly detection: A survey. *ACM Computing Surveys* **41** (2009), 15:1-15:58.
[10] Joinpoint Help Manual 4.4.0.0. Retrieved March, 2017, from https://surveillance.cancer.gov/help/joinpoint/.
[11] Joinpoint Regression Program, Version 4.4.0.0. (2017) Statistical Methodology and Applications Branch, National Cancer Institute. Retrieved January, 2017, from https://surveillance.cancer.gov/.
[12] Julious, S. A.: Inference and estimation in a changepoint regression problem. *Journal of the Royal Statistical Society. Series D (The Statistician)* **50** (2001), 51-61.
[13] Killick, R., and Eckley, I.A.: changepoint: An R package for changepoint analysis. *Journal of Statistical Software* **58** (2014). doi: 10.18637/jss.v058.i03.
[14] Kim, H. J., Feuer, E. J. and Midhune, D. N.: Permutation Tests for Joinpoint Regression with Applications to Cancer Rates. (correction: 2001; 20:655). *Statistics in Medicine* **19** (2000), 335-351.
[15] Kim, H. J., Yu, B. and Feuer, E. J.: Selecting the number of change-points in segmented line regression. *Statistica Sinica* **19** (2009), 597-609.
[16] Kim, J. and Kim, H. J.: Asymptotic results in segmented multiple regression. *Journal of Multivariate Analysis* **99** (2008), 2016-2038.
[17] Maciak, M. and Mizera, M.: Regularization Techniques in joinpoint regression. *Stat. Papers* **57** (2016), 939-955.
[18] Maidstone, R., Hocking, T., and Rigail, G.: On optimal multiple changepoint algorithms for large data. *Stat. Comput.* **27** (2017), 519-533.
[19] Milborrow, S.: earth: Multivariate Adaptive Regression Splines. (2016). Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller's Fortran utilities with Thomas Lumley's leaps wrapper. Retrieved Feb., 2017, from https://cran.r-project.org/web/packages/earth/index.html.
[20] Milborrow, S.: Notes on the earth package. Retrieved Feb., 2017, from http://www.milbo.org/doc/earth-notes.pdf.
[21] Muggeo, V. M.: Estimating regression models with unknown break-points. *Statistics in Medicine* **22** (2003), 3055-3071.
[22] Muggeo, V. M.: segmented: An R package to fit regression models with broken-line relationships. *R News* **8** (2008), 20-25.
[23] Muggeo, V. M.: Package segmented ver. 0.5 - 1.4. (2015). Retrieved Feb., 2017, from https://cran.r-project.org/web/packages/ segmented/segmented.pdf.
[24] R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2014. Retrieved January, 2017, from http://www.R-project.org/.
[25] Ross, G. J.: Parametric and Nonparametric Sequential Change Detection in R: The cpm Package. *Journal of Statistical Software* **66** (2015). doi: 10.18637/jss.v066.i03.
[26] Zhang, H., & Singer, B.: *Recursive partitioning in the health sciences*. Springer-Verlag Inc., New York 1999.

Timetable Construction for a Village Small School

Veronika Skocdopolova[1], Mirka Simonovska[2]

Abstract. Small schools form a considerable part of all primary schools in the Czech Republic. Small school is a school where children from different grades are taught in at least one class. We can find this type of school especially in small villages, where is not enough children for regular primary school.
The paper presents computational approach for timetable construction at a small school in a village in the Central Bohemian Region. On the timetable for this type of primary schools there are imposed greater demands than on other classes, that allow to teach every class on the first stage separately. In particular, it must be ensured that the classes with joint grades are taught the same subject in as many lessons as possible. For the timetable construction, an integer programming model based on assignment problem is formulated. The aim of the model is to minimise the number of afternoon lessons with respect to restrictions given by the School Education Programme.

Keywords: small school, timetabling, integer programming, assignment problem.

JEL Classification: C61
AMS Classification: 90C10

1 Introduction

Timetabling is a way, how to construct not only school timetable effectively [2]. It helps avoiding time or resource wasting and therefore it reduces cost. There are two main approaches to timetable construction using mathematical modelling. The first one is creating a complex model usually using binary variables. The other one is based on a decomposition of the problem into several interrelated stages. [4] More about timetabling can be found in [1] or [3].

This paper handle a problem of timetable construction at the primary school in a small village Višňová in the Central Bohemian Region. This primary school is a small school, in which pupils from different grades are taught in joint classes. This fact has to be taken into consideration during the timetable construction.

2 Small School

The small school is a school, in which there is at least one class, where children from different grades are taught together [5]. This type of school can be found usually in small villages. Historically, there are two main reasons for a small school. The first reason is that there is lack of qualified teachers able to teach the pupils. Nearly a hundred pupils were taught in one class in that case in the past. Nowadays, this reason is significant in the less developed countries. The second reason is that there are not enough children to open a regular primary school in the region. This is specific for developed countries. Both reasons were relevant over the years in the Czech Republic.

The standard elementary school in the Czech Republic provides education in 9 grades divided into primary school (grades 1–5) and secondary school (grades 6–9). According to the Education Act No. 561/2004 Coll., it is possible to run a school that does not provide all 9 grades. Therefore a primary school can exist separately. According to the Regulation No. 48/2005 Coll., only the grades at the primary school can be joint. There are 1 113 small schools in the Czech Republic today. The small schools form 27 % of all the elementary schools and they are attended by 4.4 % of all the pupils [6].

Many pros and cons relate to small schools. However, whether a thing is an advantage or disadvantage depends on the point of view. The friendly environment in the class is a clear advantage. Most of the pupils know each other not only from the school, because they live mostly in the same village. There is a closer relationship between the teacher and the pupil in contrast to big schools. Another good point is that the small school is the only school in the village. Therefore it is supported by the municipality not only financially but also by the possibility to use local premises for school events. An important benefit results from the fact that younger and older pupils are mixed together in one class. The younger children have the opportunity to learn from the older ones. The older pupils

[1] University of Economics, Prague, department of econometrics, W. Churchill sq. 4, Prague, Czech Republic, veronika.skocdopolova@vse.cz.
[2] University of Economics, Prague, W. Churchill sq. 4, Prague, Czech Republic, xsimm90@vse.cz.

should respect the younger pupils. They learn how to make compromise, how to respect each other. Their communication skills improve. Last but not least, most of the small school pupils live in the same village hence they do not have to get up early in the morning and commute to a distant city.

On the other hand, we can find some negative points too. The main negative aspect is that teaching in a class with pupils from different grades put high demands on teachers. They have to prepare their lessons carefully. A distinctive drawback is the necessity of changing the school after finishing the fifth grade of the primary school. The pupils have to commute to a distant city and incorporate in a new group of children at the regular elementary school.

3 Problem Description and Model Formulation

The primary school in Višňová is a small school with three classes in the school year 2015/2016. 53 pupils attended the school that school year. The basic frame for the timetable is given by the School Education Programme that every school has to prepare according to the Framework Education Programme for Elementary Education. These programs define the time allocation of each subject.

3.1 Timetable Requirements

There were two joint classes in the school year 2015/2016. Which grades might be joint in one class depends on the number of pupils in each grade and on the difficulty of teaching in the grade. There were 16 pupils in the first grade in that school year. Therefore the first grade was taught separately. The second and the fifth grade were joint the previous year and there was no reason to separate them. For this reason, the third and the fourth grade were joint for that year. Hence the task is to prepare the timetable for such a joint structure.

For primary school there is a rule that there can be only four or five lessons in the morning. A lunch break has to follow after these lessons. The pupils from different grades are joint for all subjects except P.E., where the maturity of pupils plays its role. For P.E. pupils from the first, the second and the third or the fourth and the fifth grade can only be joint. Each class has its class master who teach all the subjects in the class except English, Music and Art. There is only one teacher with the teaching qualification for each of these subjects.

All the subjects should be distributed evenly over the week with respect to the time allocation of each subject. A free period must not occur during the morning lessons (before the lunch break). The difficult subjects (Czech and Maths) should be taught within the first three lessons of each day. The aim of the timetabling process is to fulfil all the requirements mentioned above and minimise the number of afternoon lessons.

3.2 Mathematical Model

The model operates with four index sets: i – grade of the primary school, $i = 1, 2, \ldots, 5$; j – day of week, $j = 1, 2, \ldots, 5$; k – time window within each day, $k = 1, 2, \ldots, 7$; and l – subjects including lunch break, $l = 1, 2, \ldots, 16$. There are 10 different subjects in the School Education Programme of the Višňová Primary School. We have to divide the subjects with different time allocation in each grade into two separate subjects for the purpose of the model formulation. The index set l is divided into several subsets. The list of all subjects and subsets, in which the subjects are included, is in the **Table 1**. The binary decision variable x_{ijkl} equals 1, if the pupils from the i-th grade on the day j in the time window k have the subject l, and 0 otherwise. We formulated the following mathematical model for the timetable construction:

The model objective is to minimise the number of afternoon lessons:

$$z = \sum_{i \in grade} \sum_{j \in day} \sum_{l \in sub} (x_{ij6l} + 2x_{ij7l}) \tag{1}$$

If there is an afternoon lesson, it should be directly after the lunch break. Therefore the value of the variable x_{ij7l} is multiplied by two.

To avoid time conflicts, there should be at most one subject in each time window for each grade:

$$\sum_{l \in sub \cup lu} x_{ijkl} \leq 1, \quad \forall i, j, k \tag{2}$$

Each subject in each grade has to correspond with its time allocation ta_{il}:

$$\sum_{j \in day} \sum_{k \in time} x_{ijkl} = ta_{il}, \quad \forall i, l \tag{3}$$

The following constraints (4) – (9) assure the even distribution of the subjects over the week. There should be at least one and at most two lessons of Czech every day in each grade:

$$1 \leq \sum_{k \in time} \sum_{l \in cz} x_{ijkl} \leq 2, \quad \forall i, j \tag{4}$$

There should be at most one lesson of each of the other subjects every day in each grade.

$$\sum_{k \in time} \sum_{l \in en} x_{ijkl} \leq 1, \quad \forall i, j \tag{5}$$

$$\sum_{k \in time} \sum_{l \in ma} x_{ijkl} \leq 1, \quad \forall i, j \tag{6}$$

$$\sum_{k \in time} \sum_{l \in sc} x_{ijkl} \leq 1, \quad \forall i, j \tag{7}$$

$$\sum_{k \in time} \sum_{l \in ar} x_{ijkl} \leq 1, \quad \forall i, j \tag{8}$$

$$\sum_{k \in time} x_{ijkl} \leq 1, \quad \forall i, j, l \in os \tag{9}$$

This constraint assures the joining of the third and the fourth grade for the subjects from the subset *js1*:

$$x_{3jkl} = x_{4jkl}, \quad \forall j, k \in am, l \in js1 \tag{10}$$

The subjects from the subset *js2* should be taught in the joint classes of the second and the fifth grade:

$$x_{2jkl} = x_{5jkl}, \quad \forall j, k \in am, l \in js2 \tag{11}$$

There should be at most one lesson of English, Art, Music and P.E. in each time window, while there is only one teacher with necessary teaching qualification (in case of English, Art and Music) and only one gym (in case of P.E.); hence the lesson can be attended by two joint grades, the left-hand side value is 2:

$$\sum_{i \in grade} \sum_{l \in en} x_{ijkl} \leq 2, \quad \forall\, j, k \tag{12}$$

$$\sum_{i \in grade} \sum_{l \in ar} x_{ijkl} \leq 2, \quad \forall\, j, k \tag{13}$$

$$\sum_{i \in grade} x_{ijkl} \leq 2, \quad \forall j, k, l \in mu \cup pe \tag{14}$$

During the lunch time (lt = 5, 6) the lunch break should take place:

$$\sum_{k \in lt} x_{ijkl} = 1, \quad \forall i, j, l \in lu \tag{15}$$

There should not be a free period in the morning (am = 1, 2, …, 5):

$$\sum_{k \in am} \sum_{l \in sub \cup lu} x_{ijkl} = 5, \quad \forall i, j \tag{16}$$

The following constraints (17) – (22) assure the correct joining of the P.E. lessons:

$$x_{1jkl} + x_{4jkl} \leq 1, \quad \forall j, k, l \in pe \tag{17}$$

$$x_{1jkl} + x_{5jkl} \leq 1, \quad \forall j, k, l \in pe \tag{18}$$

$$x_{2jkl} + x_{4jkl} \leq 1, \quad \forall j, k, l \in pe \tag{19}$$

$$x_{2jkl} + x_{5jkl} \leq 1, \quad \forall j, k, l \in pe \tag{20}$$

$$x_{3jkl} + x_{4jkl} \leq 1, \quad \forall j, k, l \in pe \tag{21}$$

$$x_{3jkl} + x_{5jkl} \leq 1, \quad \forall j, k, l \in pe \tag{22}$$

There should not be any lesson on the Friday afternoon:

$$x_{ijkl} = 0, \quad \forall i, j = 5, k = 7, \forall l \tag{23}$$

The difficult subjects that need more concentration (Czech and Maths) must not be scheduled to the fourth and later time windows (pm = 4, 5, 6, 7):

$$\sum_{k \in pm} x_{ijkl} = 0, \quad \forall i, j, l \in di \tag{24}$$

The decision variable x_{ijkl} is a binary variable:

$$x_{ijkl} \in \{0,1\}, \quad \forall i, j, k, l \tag{25}$$

index	subject	included in subsets
1	Czech	*sub, cz, js1, js2, di*
2	English	*sub, en, js1, js2*
3	Math	*sub, ma, js1, js2, di*
4	Informatics	*sub, os, js1*
5	Science	*sub, sc, js1, js2*
6	Music	*sub, os, js1, js2*
7	Art	*sub, ar, mu, s1, js2*
8	Drama	*sub, os, js1, js2*
9	P. E.	*sub, os, pe*
10	Working activities	*sub, os, js1, js2*
11	Czech 2	*sub, cz*
12	English 2	*sub, en*
13	Math 2	*sub, ma*
14	Science 2	*sub, sc*
15	Art 2	*sub, ar*
16	Lunch break	*lu*

Table 1 Subjects taught in the school

4 Results and Conclusion

The result of our model is a timetable for each grade. The timetables respect the requirements and the afternoon classes have only the fourth and the fifth grade. In most of the joint lessons the same subjects for both grades are taught. Subjects Maths and Czech are assigned to the morning lessons. The model was solved using LINGO 16.0 solver. The input data were organised in MS Excel worksheet. The Figure 1 Timetable for the 3rd and 4th grade shows the timetable for the joint second and fifth grade.

		Lesson						
Day	**Grade**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Monday	3rd	Czech	Math	Czech	Art	Science	Lunch	
	4th							
Tuesday	3rd	Drama	Math	Czech	English	P. E.	Lunch	
	4th					Science 2		P. E.
Wednesday	3rd	English	Math	Czech	Science	Czech 2	Lunch	
	4th					P. E.		
Thursday	3rd	Czech	Czech	Math	P. E.	Working activities	Lunch	
	4th				Art 2			
Friday	3rd	Czech	Music	Math	English	P. E.	Lunch	
	4th					Science 2		

Figure 1 Timetable for the 3rd and 4th grade

The timetable of teachers as well as number of classes for one teacher is not considered, because the teachers are class masters. Therefore their timetable corresponds to the timetable of the class – five lessons in the morning, then lunch and someday afternoon classes (except Friday).

The calculated timetables are quite close to the real timetables designed by the school director. Thus the time-table could be prepared automatically using the proposed model. The only thing that varies each school year is which grades should be joint. This means, that only the constraints (10) and (11) and subject subsets *js1* and *js2* have to be modified every year. Therefore a simple application utilising the cooperation between MS Excel and LINGO solver can be prepared for ordinary use in the school.

Acknowledgements

The research is supported by the Internal Grant Agency of the University of Economics, Prague, grant No. F4/58/2017.

References

[1] Asratian, A. S., Werra, D.: A generalized class teacher model for some timetabling problems. *European Journal of Operational Research* **143** (2002), 531–542.

[2] Birbas, T., Daskalaki, S., Housos, E.: School timetabling for quality student and teacher schedules. *Journal of Scheduling* **12** (2009), 177–197.

[3] Pillay, N.: A survey of school timetabling research. *Annals of Operations Research* **218** (2014), 261–293.

[4] Skocdopolova, V.: A Comparison of Integer Goal Programming Models for Timetabling. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics 2016* (Kocourek, A. and Vavroušek, M., eds.). Technical University of Liberec, Liberec, 2016, 768–772.

[5] Trnkova, K., Knotova, D., Chaloupkova, L.: *Small Schools in the Czech Republic* (in Czech). Paido, Brno, 2010.

[6] Trojan, V.: Místo malotřídních škol ve vzdělávacím systému. In: *Malotřídky* 2016. Prague, 2016. Retrieved from https://slideslive.com/38896592/misto-malotridnich-skol-ve-vzdelavacim-systemu.

Risk-Sensitive Optimality in Markov Games

Karel Sladký[1], Victor Manuel Martínez Cortés[2]

Abstract. The article is devoted to risk-sensitive optimality in Markov games. Attention is focused on Markov games evolving on communicating Markov chains with two-players with opposite aims. Considering risk-sensitive optimality criteria means that total reward generated by the game is evaluated by exponential utility function with a given risk-sensitive coefficient. In particular, the first player (resp. the second player) tries to maximize (resp. minimize) the long-run risk-sensitive average reward. Observe that if the second player is dummy, the problem is reduced to finding optimal policy of the Markov decision chain with the risk-sensitive optimality. Recall that for the risk sensitivity coefficient equal to zero we arrive at traditional optimality criteria. In this article, connections between risk-sensitive and risk-neutral Markov decision chains and Markov games models are studied using discrepancy functions. Explicit formulae for bounds on the risk-sensitive average long-run reward are reported. Policy iteration algorithm for finding suboptimal policies of both players is suggested. The obtained results are illustrated on numerical example.

Keywords: dynamic programming, Markov decision chains, two-person Markov games, communicating Markov chains, risk-sensitive optimality.

JEL classification: C44, C61, C63
AMS classification: 90C40, 60J10, 93E20

1 Introduction

This contribution is devoted to risk-sensitive optimality in Markov games evolving on communicating Markov chains with two-players with opposite aims. In particular, the first player (resp. the second player) tries to maximize (resp. minimize) the long-run risk-sensitive average reward calculated by an exponential utility function with a given risk-sensitive coefficient. Observe that if the second player is dummy, the problem is reduced to finding optimal policy of the risk-sensitive Markov decision chain introduced by Howard and Matheson in their seminal paper [8]. Recall that for the risk sensitivity coefficient equal to zero we arrive at traditional optimality criteria. In this article, connections between risk-sensitive and risk-neutral Markov decision chains and Markov games models are studied using discrepancy functions. Explicit formulae for bounds on the risk-sensitive average long-run reward are reported. Policy iteration algorithms for finding suboptimal policies of both players are suggested.

2 Notation and Preliminaries

In this note, we consider at discrete time points $t = 0, 1, \ldots$ a dynamic system $X = \{X_n, n = 0, 1, \ldots\}$ with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$. The behavior of the system X is influenced by two players, $P^{(1)}$ and $P^{(2)}$, with opposite aims. Supposing that at time t the system is in state $i \in \mathcal{I}$ then player $P^{(1)}$, resp. player $P^{(2)}$, selects action $a^{(1)}$ from finite set $\mathcal{A}_i^{(1)}$, resp. action $a^{(2)}$ from finite set $\mathcal{A}_i^{(2)}$. Then state j is reached in the next transition with a given probability $p_{ij}(a^{(1)}, a^{(2)})$ and one-stage reward $r_i(a^{(1)}, a^{(2)})$ is accrued. We shall call this two person game a Markov game.

In this note, we assume that the stream of rewards generated by the Markov processes X is evaluated by an exponential utility function (so-called risk-sensitive models) with a given risk sensitivity coefficient. To this end, let us consider an exponential utility function, say $\bar{u}^\gamma(\cdot)$, i.e. a separable utility function with constant risk sensitivity $\gamma \in \mathbb{R}$. Then the utility assigned to the (random) outcome ξ is given by

$$\bar{u}^\gamma(\xi) := \begin{cases} (\text{sign } \gamma)\ \exp(\gamma\xi), & \text{if } \gamma \neq 0, \quad \text{risk-sensitive case,} \\ \xi \qquad \text{for } \gamma = 0 & \text{risk-neutral case.} \end{cases} \tag{1}$$

Obviously $\bar{u}^\gamma(\cdot)$ is continuous and strictly increasing. For $\gamma > 0$ $\bar{u}^\gamma(\cdot)$ is convex, if $\gamma < 0$ $\bar{u}^\gamma(\cdot)$ is concave. Finally if $\gamma = 0$ (risk neutral case) $\bar{u}^\gamma(\cdot)$ is linear. Observe that exponential utility function $\bar{u}^\gamma(\cdot)$ is separable and

[1]Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, sladky@utia.cas.cz
[2]Department of Mathematics, Autonomous Metropolitan University, Iztapalapa Campus, Mexico, mat.victor.m.mtz.c@gmail.com

multiplicative if the risk sensitivity $\gamma \neq 0$ and additive for $\gamma = 0$. In particular, for $u^\gamma(\cdot) := \exp(\gamma\xi)$ we have $u^\gamma(\xi_1 + \xi_2) = u^\gamma(\xi_1) \cdot u^\gamma(\xi_2)$ if $\gamma \neq 0$ and $u^\gamma(\xi_1 + \xi_2) \equiv \xi_1 + \xi_2$ for $\gamma = 0$.

Moreover, recall that the certainty equivalent corresponding to ξ, say $Z^\gamma(\xi)$, is given by

$$\bar{u}^\gamma(Z^\gamma(\xi)) = \mathsf{E}\,[\bar{u}^\gamma(\xi)] \qquad \text{(the symbol } \mathsf{E} \text{ is reserved for expectation).} \tag{2}$$

From (1), (2) we can immediately conclude that

$$Z^\gamma(\xi) = \begin{cases} \gamma^{-1}\,\ln\{\mathsf{E}\; u^\gamma(\xi)\}, & \text{if } \gamma \neq 0 \\ \mathsf{E}\,[\xi] & \text{for } \gamma = 0. \end{cases} \tag{3}$$

The development of the system X over time is controlled by actions of both players that have complete information about the history of the system. In particular, player $P^{(1)}$, resp. player $P^{(2)}$, tries to maximize, resp. minimizes the total reward. Supposing that the system is in state $i \in \mathcal{I}$ if decision $a^{(1)} \in \mathcal{A}_i^{(1)}$ is taken by the first player, player $P^{(2)}$ selects decision $a^{(2)} \in \mathcal{A}_i^{(2)}$ such that to "minimize" possible outcome (so decisions $a^{(1)}, a^{(2)}$ are not simultaneous, player $P^{(1)}$ is the leader and player $P^{(2)}$ is the follower in the considered Stackelberg duopoly model). Risk-sensitive Markov decision chains can be considered as a special case of Markov games with only one player.

A (Markovian) policy controlling the decision process, $\pi = (f^0, f^1, \ldots)$, is identified by a sequence of decision vectors $\{f^n, n = 0, 1, \ldots\}$ where $f^n = (f^{(1)n}, f^{(2)n}) \in \mathcal{F} \equiv \mathcal{F}^{(1)} \times \mathcal{F}^{(2)}$. In particular, player $P^{(1)}$, resp. player $P^{(2)}$, generates a sequence of decisions $f^{(1),n}$ where $f^{(1),n} \in \mathcal{F}^{(1)} \equiv \mathcal{A}_1^{(1)} \times \ldots \times \mathcal{A}_N^{(1)}$, resp. $f^{(2),n}$ where $f^{(2),n} \in \mathcal{F}^{(2)} \equiv \mathcal{A}_1^{(2)} \times \ldots \times \mathcal{A}_N^{(2)}$.

Let $\pi^m = (f^m, f^{m+1}, \ldots)$, hence $\pi = (f^0, f^1, \ldots, f^{m-1}, \pi^m)$, in particular $\pi = (f^0, \pi^1)$. The symbol E_i^π denotes the expectation if $X_0 = i$ and policy $\pi = (f^n)$ is followed, in particular, $\mathsf{E}_i^\pi(X_m = j) = \sum_{i_j \in \mathcal{I}} p_{i,i_1}(f_i^0) \ldots p_{i_{m-1},j}(f_{m-1}^{m-1})$; $\mathrm{P}(X_m = j)$ is the probability that X is in state j at time m.

Policy π which selects at all times the same decision rule, i.e. $\pi \sim (f)$, is called stationary. Hence following policy $\pi \sim (f)$ X is a homogeneous Markov chain with transition probability matrix $P(f)$ whose ij-th element is $p_{ij}(f_i) = p_{ij}(f_i^{(1)}, f_i^{(2)})$. Then $r_i(f_i) := r_i(f_i^{(1)}, f_i^{(2)})$ is the one-stage reward obtained in state i. Similarly, $r(f)$ is an N-column vector of one-stage rewards whose i-the elements equals $r_i(f_i)$.

Stationary policy $\tilde{\pi}$ is randomized if there exist decision vectors $f^{[1]},\ f^{[2]}, \ldots, f^{[m]} \in \mathcal{F}$ (observe that $f^{[1]} = (f^{[1](1)}, f^{[1](2)}) \in \mathcal{F}^{(1)} \times \mathcal{F}^{(2)}$). On following policy $\tilde{\pi}$ we select in state i action $f_i^{[j]}$ with a given probability $\kappa_i^{[j]}$ (of course, $\kappa_i^{[j]} \geq 0$ with $\sum_j \kappa_i^{(j)} = 1$ for all $i \in \mathcal{I}$). Observe that $\mathsf{E}_i^\pi(X_m = j) = [P^m(f)]_{ij}$ (here $[A]_{ij}$ denotes the ij-th element of the matrix A, $A \geq B$, resp. $A > B$ iff for each i, j $[A]_{ij} \geq [B]_{ij}$ resp. $[A]_{ij} \geq [B]_{ij}$ and $[A]_{ij} > [B]_{ij}$ for some i, j). The symbol I denotes an identity matrix and e is reserved for a unit column vector.

3 Risk-Sensitive Optimality in Markov Processes

Let ξ_n be the cumulative reward obtained in the n first transition of the considered Markov chain X. Since the process starts in state X_0, $\xi_n = \sum_{k=0}^{n-1} r_{X_k}$. Similarly let $\xi_{(m,n)}$ be reserved for the cumulative (random) reward, obtained from the mth up to the nth transition (obviously, $\xi_n = r_{X_0} + \xi_{(1,n)}$, we tacitly assume that $\xi_{(1,n)}$ starts in state X_1).

On introducing for arbitrary $g, w_j \in \mathbb{R}$ $(i, j \in \mathcal{I})$ the discrepancy function (cf. [10])
$\tilde{\varphi}_{i,j}(w, g) := r_i - w_i + w_j - g$ we can easily verify the following identity:

$$\xi_n \;\; = \;\; ng + w_{X_0} - w_{X_n} + \sum_{k=0}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w, g). \tag{4}$$

Considering the risk-sensitive models in virtue of (1), (4) for the expectation of ξ_n in the risk-sensitive case $U_i^\pi(\gamma, n) := \mathsf{E}_i^\pi \mathrm{e}^{\gamma \sum_{k=0}^{n-1} \xi_n}$ we conclude that

$$U_i^\pi(\gamma, n) = \mathrm{e}^{\gamma[ng + w_i]} \times \mathsf{E}_i^\pi \mathrm{e}^{\gamma[\sum_{k=0}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w,g) - w_{X_n}]}. \tag{5}$$

Now observe that

$$\mathsf{E}_i^{\pi}\, \mathrm{e}^{\gamma \sum_{k=0}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w,g)} = \sum_{j \in \mathcal{I}} p_{ij}(f_i^0)\, \mathrm{e}^{\gamma[r_i - w_i + w_j - g]} \times \mathsf{E}_j^{\pi^1}\, \mathrm{e}^{\gamma \sum_{k=1}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w,g)} \tag{6}$$

$$\mathsf{E}_j^{\pi}\, \{\mathrm{e}^{\gamma \sum_{k=m}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w,g)} | X_m = j\} = \sum_{\ell \in \mathcal{I}} p_{j,\ell}(f_j^m)\, \mathrm{e}^{\gamma[r_j - w_j + w_\ell - g]} \times \mathsf{E}_\ell^{\pi^{m+1}}\, \mathrm{e}^{\gamma \sum_{k=m+1}^{n-1} \tilde{\varphi}_{X_k, X_{k+1}}(w,g)}. \tag{7}$$

If stationary policy $\pi \sim (f)$ is followed (5) can be drastically simplified if the numbers g, w_j's are selected such that $\sum_{j \in \mathcal{I}} p_{ij}(f_i)\, \mathrm{e}^{\gamma \tilde{\varphi}_{ij}(g,w)} = 1$ for all $i \in \mathcal{I}$.[3] Obviously, this condition is equivalent to the following set of linear equations

$$\mathrm{e}^{\gamma[g(f) + w_i(f)]} = \sum_{j \in \mathcal{I}} p_{ij}(f_i)\ \mathrm{e}^{\gamma[r_i(f_i) + w_j(f)]} \quad (i \in \mathcal{I}) \tag{8}$$

for the values $g(f)$, $w_i(f) (i = 1, \ldots, N)$; observe that these values depend on the selected risk sensitivity coefficient γ. Eqs. (5) can be called the *γ-average reward/cost optimality equation*.

On introducing the new variables $v_i(f) := \mathrm{e}^{\gamma w_i(f)}$, $\rho(f) := \mathrm{e}^{\gamma g(f)}$, and on replacing transition probabilities $p_{ij}(f_i)$'s by general nonnegative numbers defined by $q_{ij}(f_i) := p_{ij}(f_i) \cdot \mathrm{e}^{\gamma r_i(f_i)}$ (8) can be alternatively written as the following set of equations

$$\rho(f) v_i(f) = \sum_{j \in \mathcal{I}} q_{ij}(f_i)\, v_j(f) \quad (i \in \mathcal{I}) \tag{9}$$

For what follows it is convenient to consider (9) in matrix form. To this end, we introduce (cf. [6]) $N \times N$ non-negative matrix $Q(f) = [q_{ij}(f_i)]$ with spectral radius (Perron eigenvalue) $\rho(f)$ along with its right Perron eigenvector $v(f) = [v_i(f_i)]$. Then (9) can be written in matrix form as

$$\rho(f) v(f) = Q(f) v(f). \tag{10}$$

Furthermore, if the transition probability matrix $P(f)$ is *irreducible* then also $Q(f)$ is *irreducible* and the right Perron eigenvector $v(f)$ can be selected *strictly positive.*

From (3),(5),(6),(8) we immediately get for stationary policy $\pi \sim (f)$ that

$$U_i^{\pi}(\gamma, n) = \mathrm{e}^{\gamma[n g(f) + w_i(f)]} \times \mathsf{E}_i^{\pi} \mathrm{e}^{\gamma w_{X_n}(f)}, \quad Z_i^{\pi}(\gamma, n) = \frac{1}{\gamma} \ln U_i^{\pi}(\gamma, n).$$

Hence

$$n^{-1} Z_i^{\pi}(\gamma, n) = g(f) + o(n) \tag{11}$$

(recall that $g(f) = \gamma^{-1} \ln \rho(f)$, $w_i(f) = \gamma^{-1} \ln v_i(f)$).

If the Markov chain is irreducible there exist $\hat{f}, f^* \in \mathcal{F}$ along with numbers $\hat{\rho} = \rho(\hat{f})$, $\rho^* = \rho(f^*)$ and strictly positive vectors $\hat{v} = v(\hat{f})$, with elements $v_i(\hat{f})$, and $v^* = v(f^*)$, with elements $v_i(f^*)$, such that for any $f \in \mathcal{F}$ (vectorial max and min should be considered componentwise)

$$Q(f) \cdot \hat{v} \geq \min_{f \in \mathcal{F}} \{Q(f) \cdot \hat{v}\} = Q(\hat{f}) \cdot \hat{v} = \hat{\rho} \cdot \hat{v} \tag{12}$$

$$Q(f) \cdot v^* \leq \max_{f \in \mathcal{F}} \{Q(f) \cdot v^*\} = Q(f^*) \cdot v^* = \rho^* \cdot v^* \tag{13}$$

$$\rho(\hat{f}) \equiv \hat{\rho} \leq \rho(f) \leq \rho(f^*) \equiv \rho^* \qquad \text{for all } f \in \mathcal{F}. \tag{14}$$

In words:

$\hat{\rho} \equiv \rho(\hat{f})$ (resp. $\rho^* = \rho(f^*)$) is the minimum (resp. maximum) possible eigenvalue of $Q(f)$ over all $f \in \mathcal{F}$ (cf. [1],[3],[8]).

Minimal (resp. maximal) risk-sensitive average reward $g(\hat{f}) = \gamma^{-1} \ln \rho(\hat{f})$ (resp. $g(f^*) = \gamma^{-1} \ln \rho(f^*)$).

[3] To verify this claim it suffices to apply successively (7) backwards starting time point $n - 1$ (cf. [8]).

4 Risk-Sensitive Optimality in Markov Games

In contrast to Markov decision model considered in section 3 we assume that the expected utility $U_i^{\pi}(\gamma, n)$ depends on decision $f^{(1),n}$, $f^{(2),n}$ taken by the both players. Since Markov decision processes can be considered as a very special case of Markov games, it is interesting to mention that stochastic games were formulated by Shapley [12] in 1953, many years before outburst of systematic interest in Markov decision processes. For the early results on Markov decision processes see Bellman's papers [1], [2], Bellman's monograph [3], Blackwell's paper [4] and especially Howard's book [7].

In contrast to Markov decision processes we must take into consideration decision taken by both players. Hence the optimality equations (12), (13) must be replaced by the Nash equilibrium condition, see [11]. To this end we assume that there exists at least one, in general randomized, stationary policy $\pi^* \sim (f^*)$ using in state $i \in \mathcal{I}$ (randomized) decision f_i^* selecting action $f_i^{[j]} \in \mathcal{F}_i$ with a given probability $\kappa_i^{[j]}$ (of course, $\kappa_i^{[j]} \geq 0$ with $\sum_j \kappa_i^{(j)} = 1$) depend on the policy selected). Then for any $f_i^{(1)} \in \mathcal{F}_i^{(1)}$ and any $f_i^{(2)} \in \mathcal{F}_i^{(2)}$ for the resulting decisions $f_i^d = (f_i^{(1)}, f_i^{(2)*})$, $f_i^u = (f_i^{(1)*}, f_i^{(2)})$ it holds

$$\sum_{j\in\mathcal{I}} q_{ij}(f_i^d)v_j^* \leq \sum_{j\in\mathcal{I}} q_{ij}(f_i^*)v_j^* = \rho(f^*)v_i^* \leq \sum_{j\in\mathcal{I}} q_{ij}(f_i^u)v_j^*, \tag{15}$$

or in matrix notations $\rho(f)v(f) = Q(f)v(f)$ we are looking for $f^* = (f^{(1)*}, f^{(2)*}) \in \mathcal{F}^{(1)} \times \mathcal{F}^{(2)}$ such that

$$Q(f^d)v(f^*) \leq \rho(f^*)v(f^*) = Q(f^*)v(f^*) \leq Q(f^u)v(f^*) \tag{16}$$

where $\rho(f^*) = \mathrm{e}^{\gamma g(f^*)}$, $v_i(f^*) = \mathrm{e}^{\gamma w_i(f^*)}$.

From (3),(5),(6),(8) we immediately get for stationary policy $\pi^* \sim (f^*)$ that

$$U_i^{\pi^*}(\gamma, n) = \mathrm{e}^{\gamma[ng(f^*)+w_i(f^*)]} \times \mathsf{E}_i^{\pi^*} \mathrm{e}^{\gamma w_{X_n}(f^*)}, \quad Z_i^{\pi^*}(\gamma, n) = \frac{1}{\gamma} \ln U_i^{\pi^*}(\gamma, n), \tag{17}$$

$$n^{-1} Z_i^{\pi^*}(\gamma, n) = g(f^*) + o(n). \tag{18}$$

Since the average risk-sensitive reward $g(f) = \gamma^{-1} \ln[\rho(f)]$ and $\rho(f)$ is the Perron eigenvalue of a nonnegative matrix $Q(f)$, it is well-known (see e.g. [6]) that for any $f', f'' \in \mathcal{F}$ $Q(f') \leq Q(f'') \Rightarrow \rho(f') \leq \rho(f'')$. To generate lower and upper bounds on minimal and maximal Perron eigenvalue $\rho(f^*)$ we replace elements $q_{ij}(f_i^{(1)}, f_i^{(2)})$ by their minimal and maximal possible values q'_{ij} and q''_{ij}. Then the problem is approximated by a (uncontrollable) risk-sensitive Markov chain and it is possible to generate lower and upper bounds on $\rho(f^*) = \mathrm{e}^{\gamma g(f^*)}$ by calculating Perron eigenvalues (i.e. the spectral radii) of nonnegative matrices. Unfortunately, using this approach we can expect only very rough bounds on the optimal value of the average risk-sensitive reward.

More friendly bounds can be obtained by a more detailed analysis of the set of all admissible matrices. Of course, it is reasonable to suggest algorithmic procedures that need not evaluate all admissible matrices. Algorithm 1 is a slight modification of the policy iteration method reported in [8] only for finding maximum Perron eigenvalue in a set of nonnegative irreducible matrices.

Algorithm 1. (Policy iterations for finding maximal, resp. minimal, Perron eigenvalue.)

Step 0. Find matrix $Q^{(0)} := Q(f^{(1),0}, f^{(2),0})$ with $f^{(1),0} \in \mathcal{F}^{(1)}$, $f^{(2),0} \in \mathcal{F}^{(2)}$ such that the row sums are maximal (resp. minimal).

Step 1. For matrix $Q^{(k)}$ $(k = 0, 1, \ldots)$ calculate its spectral radius $\rho^{(k)}$ along with its right Perron eigenvector $v^{(k)}$.

Step 2. Construct (if possible) matrix $Q^{(k+1)} := Q(f^{(1),k+1}, f^{(2),k+1})$ with
$f^{k+1} := (f^{(1),k+1}, f^{(2),k+1})$ where $f^{(1),k+1} \in \mathcal{F}^{(1)}$, $f^{(2),k+1} \in \mathcal{F}^{(2)}$, such that

$$Q^{(k+1)} \cdot v^{(k)} \geq \rho^{(k)} v^{(k)} = Q^{(k)} \cdot v^{(k)} \quad \text{resp.} \quad Q^{(k+1)} \cdot v^{(k)} \leq \rho^{(k)} v^{(k)} = Q^{(k)} \cdot v^{(k)} \tag{19}$$

Step 3. If $Q^{(k+1)} = Q^{(k)}$ then go to Step 4, else set $k := k + 1$ and repeat Step 1.

Step 4. Set $\hat{Q} := Q^{(k+1)}$, $\hat{\rho} := \rho^{(k+1)}$, $\hat{v} := v^{(k+1)}$, $\hat{f} := f^{(k+1)}$ and stop. Then $\hat{\rho}$ is the maximal (resp. the minimal) Perron eigenvalue.

The heart of the above algorithm is the following

Policy improvement routine:
Since for the right (resp. left) Perron eigenvectors $v^{(k)}$ (resp. $z^{(k)}$) of an irreducible matrix $Q^{(k)}$ it holds $Q^{(k)} \cdot v^{(k)} = \rho^{(k)} v^{(k)}$ (resp. $z^{(k)} Q^{(k)} = \rho^{(k)} z^{(k)}$) if $\varphi^{(k+1)} := Q^{(k+1)} \cdot v^{(k)} - Q^{(k)} \cdot v^{(k)} > 0$ (resp. $\varphi^{(k+1)} < 0$) then

$$Q^{(k+1)} \cdot v^{(k+1)} - Q^{(k)} \cdot v^{(k)} = \rho^{(k+1)}[v^{(k+1)} - v^{(k)}] + [\rho^{(k+1)} - \rho^{(k)}]v^{(k)}.$$

On premultiplying the above equality by $z^{(k+1)}$ (strictly positive row vector) we arrive at

$$\rho^{(k+1)} \cdot z^{(k+1)}[v^{(k+1)} - v^{(k)}] + [\rho^{(k+1)} - \rho^{(k)}] \cdot z^{(k+1)} v^{(k)} = z^{(k+1)} Q^{(k+1)}[v^{(k+1)} - v^{(k)}] + z^{(k+1)} \varphi^{(k+1)}$$

implying that $z^{(k+1)} \varphi^{(k+1)} = [\rho^{(k+1)} - \rho^{(k)}] z^{(k+1)} v^{(k)}$.
Since $z^{(k+1)} v^{(k)} > 0$ if $z^{(k+1)} \varphi^{(k+1)} > 0$ (resp. $z^{(k+1)} \varphi^{(k+1)} < 0$) then $\rho^{(k+1)} > \rho^{(k)}$ (resp. $\rho^{(k+1)} < \rho^{(k)}$).

Illustrative example.
Let $\mathcal{I} = \{1, 2\}$, $\mathcal{A}_1^{(1)} = \mathcal{A}_1^{(2)} = \mathcal{A}_2^{(1)} = \mathcal{A}_2^{(2)} = \{1, 2\}$ and the corresponding transition probabilities be given by the row vectors $p_i(f_i^{(1)}, f_i^{(2)}) = [p_{i1}(f_i^{(1)}, f_i^{(2)}), p_{i2}(f_i^{(1)}, f_i^{(2)})]$ for $f_i^{(1)}, f_i^{(2)} = 1, 2$. The reward accrued in state i is equal to $r_i(f_i^{(1)}, f_i^{(2)})$.
The following example is borrowed from [5], Example 3.2.2, page 96.
Let transition data and one-stage rewards be:

$p_1(1,1) = [0.5; 0.5]$	$r_1(1,1) = 10$	$p_1(1,2) = [0.5; 0.5]$	$r_1(1,2) = -6$
$p_1(2,1) = [0.8; 0.2]$	$r_1(2,1) = -4$	$p_1(2,2) = [0.8; 0.2]$	$r_2(2,2) = 8$
$p_2(1,1) = [0.3; 0.7]$	$r_2(1,1) = -2$	$p_1(1,2) = [0.3; 0.7]$	$r_2(1,2) = 5$
$p_2(2,1) = [0.9; 0.1]$	$r_2(2,1) = 4$	$p_2(2,2) = [0.9; 0.1]$	$r_2(2,2) = -10$

Considering the risk-sensitive model, we replace one-stage reward $r_i(f_i^{(1)}, f_i^{(2)})$ by
$\bar{r}_i(f_i^{(1)}, f_i^{(2)}) := \ln[r_i(f_i^{(1)}, f_i^{(2)})]$ if $r_i(f_i^{(1)}, f_i^{(2)}) > 0$ or by
$\bar{r}_i(f_i^{(1)}, f_i^{(2)}) := \ln[-r_i(f_i^{(1)}, f_i^{(2)})]$ if $r_i(f_i^{(1)}, f_i^{(2)}) < 0$.
Observe that $\mathrm{e}^{\gamma \bar{r}_i(f_i^{(1)}, f_i^{(2)})} = |r_i(f_i^{(1)}, f_i^{(2)})|^\gamma$. On recalling that
$q_{ij}(f_i^{(1)}, f_i^{(2)}) =: p_{ij}(f_i^{(1)}, f_i^{(2)}) \times r_i(f_i^{(1)}, f_i^{(2)})$, let the row vectors
$q_i(f_i^{(1)}, f_i^{(2)}) =: [q_{i1}(f_i^{(1)}, f_i^{(2)}), q_{i2}(f_i^{(1)}, f_i^{(2)})]$.
Then $Q(f^{(1)}, f^{(2)})$ is the square (nonnegative) matrix whose i-th row is equal to $q_i(f_i^{(1)}, f_i^{(2)})$.

In particular, if $\gamma = 1$, resp. $\gamma = 0.5$,

$\gamma = 1$	$\gamma = 1$		$\gamma = 0.5$	$\gamma = 0.5$
$q_1(1,1) = [5; 5]$	$q_1(1,2) = [3; 3]$		$q_1(1,1) = [1.581; 1.581]$	$q_1(1,2) = [1.225; 1.225]$
$q_1(2,1) = [3.2; 0.8]$	$q_1(2,2) = [6.4; 1.6]$		$q_1(2,1) = [1.6; 0.4]$	$q_1(2,2) = [2.2628; 0.5656]$
$q_2(1,1) = [0.6; 1.4]$	$q_2(1,2) = [1.5; 3.5]$		$q_2(1,1) = [0.4242; 0.9899]$	$q_2(1,2) = [0.671; 1.5652]$
$q_2(2,1) = [3.6; 0.4]$	$q_2(2,2) = [9; 1]$		$q_2(2,1) = [1.8; 0.2]$	$q_2(2,2) = [2.846; 0.3162]$

As we can see, if $\gamma = 1$, on selecting in state 1 decision (1,1) and in state 2 decision (2,2) spectral radius of the resulting matrix is equal to 10 – maximum possible value. Similarly, selecting in state 1 decision (2,1) and in state 2 decision (1,1) spectral radius of the resulting matrix is equal to 3.4358 – minimum possible eigenvalue.
However, if $\gamma = 0.5$, on selecting in state 1 decision (1,1) and in state 2 decision (2,2) spectral radius of the resulting matrix is equal to 3.1621 – maximum possible value. Minimum possible value of the spectral radius is again obtained on selecting in state 1 decision (2,1) and in state 2 decision (1,1) spectral radius of the resulting matrix is equal to 1.8075, very close to spectral radius 1.8378 obtained if in state 1 decision (1,2) is selected and in state 2 decision (1,1) is unchanged.
Obviously, if $\gamma = 0$ the spectral radius equals one for all decisions.

Moreover, if the risk-sensitive coefficient $\gamma = 1$ by a direct calculation we can see that if the second players selects decision 2 (resp.1) in both states the first player maximize the profit by selecting action 2 in both states (resp. action 1 in state 1 and action 2 in state 2). If the second players selects decision 2 in state 1 and decision 1 in state 2 the optimal policy of the player 1 is to select action 2 in states 1 and 2. Finally, if the second players selects decision 1 in state 1 and decision 2 in state 2 the optimal policy of the player 1 is to select in state 1 action 1 and action 2 in state 2. Observe that in this case it is necessary to solve 4 problems concerning finding optimal policy of a risk-sensitive Markov decision chain.

To this end we suggest the following algorithmic procedure. More details and some numerical examples can be found in [13]. Observe that we restrict only on non-randomized decisions.

Algorithm 2. (Policy iterations for approximating optimal average reward.)

Step 0. Find matrix $Q^{(0)} := Q(f^{(1),0}, f^{(2),0})$ with $f^{(1),0} \in \mathcal{F}^{(1)}$, $f^{(2),0} \in \mathcal{F}^{(2)}$ such that its spectral radius is maximal (resp. minimal).

Step 1. For matrix $Q^{(k)}$ $(k = 0, 1, \ldots)$ calculate its spectral radius $\rho^{(k)}$ along with its right Perron eigenvector $v^{(k)}$.

Step 2. Construct (if possible) matrix $Q^{(k+1)} := Q(f^{(1),k+1}, f^{(2),k+1})$ with
$f^{k+1} := (f^{(1),k+1}, f^{(2),k+1})$ where $f^{(1),k+1} \in \mathcal{F}^{(1)}$, $f^{(2),k+1} \in \mathcal{F}^{(2)}$, such that
$f^{(1),k+1} = f^{(1),k}$ for k odd, resp. $f^{(2),k+1} = f^{(2),k}$ for k even, and

$$Q^{(k+1)} \cdot v^{(k)} \leq \rho^{(k)}\, v^{(k)} = Q^{(k)} \cdot v^{(k)} \quad \text{if } k \text{ is odd} \qquad \text{resp.} \tag{20}$$

$$Q^{(k+1)} \cdot v^{(k)} \geq \rho^{(k)}\, v^{(k)} = Q^{(k)} \cdot v^{(k)} \quad \text{if } k \text{ is even} \tag{21}$$

Step 3. If for some $\ell = 0, 1, \ldots k$ it happens that $Q^{(k+1)} = Q^{(\ell)}$ then go to Step 4, else set $k := k + 1$ and repeat Step 1.

Step 4. Set $\overline{Q} := Q^{(\ell)} Q^{(\ell+1)} \cdot \ldots Q^{(k)}$. Calculate $\bar{\rho}$, the spectral radius of $\overline{Q}$ and stop.

Then $\rho^* = (\bar{\rho})^{\frac{1}{k-\ell}}$ is equal to the long-run risk-sensitive average reward generated by decisions of the first and second player in the class of non-randomized policies.

Acknowledgements

This research was supported by the Czech Science Foundation under Grant 15-10331S and by CONACyT (Mexico) and AS CR (Czech Republic) under Project 171396.

References

[1] Bellman, R.: On a quasi-linear equation, *Canad. J. of Math.* **8** (1956), 198-202.

[2] Bellman, R.: A Markovian decision process, *J. of Math. and Mech.* **6** (1957), 679-684.

[3] Bellman, R.: *Dynamic Programming.* Princeton Univ. Press, Princeton. Princeton 1957.

[4] Blackwell, D.: Discrete dynamic programming, *Ann. Mathem. Statist.* **33** (1962), 719-726.

[5] Filar, J.A. and Vrieze, O.J.: *Competitive Markov Decision Processes.* Springer Verlag, Berlin, 1996.

[6] Gantmakher, F. R.: *The Theory of Matrices.* Chelsea, London, 1959.

[7] Howard, R. A.: *Dynamic Programming and Markov Processes.* MIT Press, Cambridge, Mass., 1960.

[8] Howard, R. A. and Matheson, J.: Risk-sensitive Markov decision processes, *Manag. Sci.* **23** (1972), 356–369.

[9] Kolokoltsov, V.N. and Malafayev, O.A.: *Understanding Game Theory.* World Scientific, Singapore, 2010.

[10] Mandl, P.: On the variance in controlled Markov chains, *Kybernetika* **7** (1971), 1–12.

[11] Nash, J.: Equilibrium points in n-person games, *Proc. Nat. Acad. Sci. U.S.A.* **36** (1950), 48–49.

[12] Shapley, L.S.: Stochastic games, *Proc. Nat. Acad. Sci. U.S.A.* **39** (1953), 1095–1100.

[13] Sladký, K. and Martínez-Cortés, V.M.: *Risk-Sensitive Optimality in Markov Games.* Research Report No.2361, Institute of Information Theory and Automation of the CAS, Prague 2017.

Use of Simulation Methods for Evaluation of Alzheimer's Disease Early Detection in Czechia

Hana Marie Smrčková[1], Václav Sládek[2], Markéta Arltová[3], Jakub Černý[4]

Abstract. The aim of our paper is to develop a model that would reliably estimate the costs of an early detection scheme for Alzheimer's disease (AD) in Czechia. In health economics, three types of models are most commonly used to this purpose: Markov cohort model, Monte Carlo simulation and discrete event simulation (DES). In the first part of the paper, we discuss the advantages and disadvantages of each method and their use in the context of the Czech health-care system and data availability. It is argued that DES is the preferred method with regards to results accuracy since the other methods rely more heavily on a number of questionable assumptions. However, the quality of the data we have collected for Czechia does not allow us to directly use this method. Consequently, we start to build our model on the methodology using Monte Carlo simulation that was previously established to estimate economic outcomes of early identification of AD in the USA. Bearing in mind the specificities of the Czech context as well as the existing data sources, and upon careful scrutiny of each parameter and assumption, we modify this model to bring it as close to DES as possible.

Keywords: Monte Carlo simulation, Markov cohort model, discrete event simulation, health economics

JEL Classification: C15, C54, I10, I18

AMS Classification: 65C05, 65C40

1 Introduction

Alzheimer's disease (AD) is the most common form of dementia that affects a significant part of elderly population. Czech Alzheimer Society estimates that 153 thousand people suffered from dementia in Czechia in 2015, out of whom approximately 60% from AD [11]. This neurodegenerative illness represents a heavy economic burden for health care systems as well as for family budgets. The high costs of care are mainly induced by its progressive character that lowers patients' ability to perform basic daily activities such as food preparation or personal hygiene. Thus, extensive assistance of family members, health or social care workers is required and substantial part of patients with more severe phases of AD are in institutional long term care. According to recent research [7], average monthly costs of care for AD patients are €1,949 – only €230 of this amount can be ascribed to treatment whereas the rest pertains to informal assistance costs. The need for assistance increases over time with the irreversible cognitive decline caused by the AD. While patients with mild cognitive impairment monthly consume about 150 hours of informal care, this figure more than doubles to 340 hours for those with severe impairment.

Although AD cannot be cured, available treatment hinders its progression and cognitive decline, especially if commenced early after the disease onset. Consequently, patients spend longer time with a milder form of dementia, which leads to better quality of life. Growing economic evidence suggests that early detection and immediate treatment represent a cost-effective strategy, potentially leading to sizable savings [5, 16]. In Czechia, no early detection scheme has been widely implemented, and majority of patients are diagnosed too late or remain undiagnosed (only about one quarter are treated at all)[11]; all the AD patients, nevertheless, consume informal care. It thus appears likely that implementation of an early diagnosis scheme would prove cost-effective similarly to other countries. This intuition is supported by a study by Mohleska et al. [12] who show on a sample of AD

[1] University of Economics, Prague, Faculty of the Informatics and Statistics, nám. Winstona Churchilla 1938/4, 130 67 Praha 3, Czech Republic (all authors), xsmrh00@vse.cz
[2] xslav20@vse.cz
[3] arltova@vse.cz
[4] xcerj129@vse.cz

patients that early diagnosis is linked to lower costs. However, no reliable statistical model estimating cost-effectiveness of AD early detection and treatment has been built for Czechia so far. Our paper addresses this gap. First, we briefly present the advantages and shortcomings of the main methodological approaches used to model cost-effectiveness for AD, namely Markov cohort model, Monte Carlo simulation and discrete event simulation. Secondly, we estimate cost-effectiveness in Czechia using Monte Carlo simulation and discuss how to modify the model to overcome its drawbacks.

2 Approaches to model AD cost-effectiveness

Generally, the structure of most models simulating (cost-)effectiveness of AD interventions is similar regardless of the chosen method. There are two branches of the model representing patient's outcomes with and without intervention (the latter being called care as usual, CAU). The disease progression is typically measured in earned points from the Mini Mental State Examination test (MMSE) or similar scale that is widely used for assessment of cognitive functions. According to her MMSE score, a patient is classified as suffering from mild (MMSE 28-21), moderate (MMSE 20-11) or severe (MMSE 10-1) dementia. The speed with which the disease progresses is modelled according to schemes of cognitive decline coming from medical literature. The three methods we discuss here – Markov cohort models (MCM), Monte Carlo simulation (MCS) and discrete event simulation (DES) – differ mainly in two aspects: the nature of modeled objects and conceptualization of the disease course [2]. In other words, we distinguish between models that consider either the entire population of patients or an individual patient as a unit of simulation. Further, one can think of a disease as about series of states where an object moves from one to another or, alternatively, as about events the object can experience. These differences carry important implications for model accuracy.

Markov cohort models are widely used in health economics modelling including studies on AD [7]. This model is used, for instance, by the rare domestic statistical AD study examining cost-effectiveness of a ginkgo biloba extract for patients with milder forms of dementia [9]. Markov models are based on a Markov property whereby transition to another state is dependent only on the current state, not on preceding states. In our opinion, this major assumption is, however, very likely violated since moving from one state to next depends on personal history. Therefore, the results of Markov models are likely biased.

Moreover, cohort modelling suffers from two additional issues: determining a relevant proportion of population that moves from one state to another at a given point in time, and application of competing risks over time [2]. Since probability of being sick depends on various personal characteristics, modelling a homogeneous population usually requires working with mean probability which represents a major simplification. Moreover, the transition probability is not constant over time in the population because high-risk groups move to the sick state first, and the probability decreases for the rest. To put it differently, MCM loses memory of time when a subgroup moves from one state to another, which results into constant transition probabilities that cannot represent disease course reliably in most cases. Solving the problem by keeping subgroups separated comes at the expense of increased model complexity [14]. Furthermore, other health risks should be taken into account because leading them leads to biased inferences. Then the matrix of transition probabilities should be multidimensional, demanding a large data set which is virtually impossible to collect. Perhaps using some a priori information (from other countries) could solve this data problem.

Monte Carlo simulation effectively eliminates the two mentioned issues since it considers individuals instead of cohorts. In this case, risk is adjusted according to personal characteristics (at least to sex and age), risk can be modified over time, and an individual can face a number of competing risks simultaneously. Consequently, MCS is more accurate and thus preferred technique to MCM albeit slightly more data-demanding. This method is used, for instance, to calculate cost-effectiveness of AD early assessment and treatment in the United States [16]. Nevertheless, MCS suffers from another major issue of Markov models: mutual exclusivity of states. Markov models, either cohort or individual, conceptualize any disease, including AD, as series of states among which patient migrates at a given point in time. In each period, an object can move one state up or down, or stay in the current one. In the context of health care, it is, however, difficult to design states that would accurately represent reality. Patients experience various medical checks, treatment strategies and changes in their health states that depend on their personal history, combination of which is important for representation of the patients' risks. Trying to include all possible combinations as a separate state (e.g. having AD and being hospitalized with a hip fracture at the same time) would lead to unreasonable increase in model complexity.

For that reason, it seems more natural to conceptualize the disease progression as a continuum of events that a patient can experience and that affects her future risks rather than a series of pre-defined states [2]. Such an approach is used by DES which, consequently, gives the most accurate results. There is no unanimous agreement about definition of differences between MCS and DES and the relationship of these methods has evolved over time. Originally, MCS was defined as a static simulation model representing a system at a particular point in time

[1]. In contrast, DES approach was more suitable for dynamic problems. However, the term MCS became gradually a synonym for stochastic computer simulation [13]. Following this logic, some forms of DES could be seen as a particular realization of MCS. Nevertheless, the difference important for the application of these methods in modelling of AD lies in distinct conceptualization of how the disease progresses in time. The fact that DES conceptualizes the disease progression as a continuum of events gives this method a considerable flexibility and allows it to model even very comprehensive treatment strategies. On the other hand, one should consider that DES is much more data-demanding than MCM or MCS. Also, DES requires more calculations and its structure seems to be too much complex at first sight. This is, however, given mainly by the fact that DES indeed considers a greater number of (important) details of the simulated case. Trying to achieve comparable accuracy by using a variant of Markov model would very likely lead to an even more complicated structure [15] and increase in number of calculations. At any rate, it is challenging to convey a DES model in a well-understandable and transparent manner. At the same time, the need to include a detailed appendix describing model structure might help transparency in the end [2]. Nowadays, some aspects and assumptions regarding standardly used models – including the original MCS model replicated in this study – are only vaguely described in the studies which hinders their critical review or use by other researchers.

3 Monte Carlo simulation for Czechia

Although DES is supposed to bring the most accurate results, individual patient data that are detailed enough to use this method are not currently available in Czechia. As it overcomes some inaccuracies of Markov cohort model, we replicate a Monte Carlo simulation model developed originally by Weimer [16] to evaluate economic outcomes of early assessment and treatment of AD patients in the United States. To be sure, we use Czech cost information and survival probability. An individual patient is simulated one thousand times in a hypothetical case that *early detection* existed. The result is compared to the simulation of *care as usual*. All simulations were programmed and run in R software. Assumptions and parameters of this model are summarized in **Table 1**.

Parameter:	**Explanation:**	**Source:**
MMSE decline model/year		
Mean decline model:		[16]
Annual decline without treatment	Normal distribution with a mean of 3.5 and a standard deviation of 1.5, with negative truncation	
Annual decline with treatment	Normal distribution with a mean of 1.5 and a standard deviation of 1.5, with negative truncation	
Decline Model of Lopez:		[10]
Probability of being slow progressor without treatment	0.39	
Annual decline of slow progressors without treatment	Uniform distribution over range of 21 to 2: mean, 0.5	
Annual decline of fast progressors without treatment	Uniform distribution over range of 3 to 6.8: mean, 4.9	
Probability of being slow progressor with treatment	0.60	
Annual decline of slow progressors without treatment	Uniform distribution over range of 21 to 2: mean, 0.5	
Annual decline of fast progressors without treatment	Uniform distribution over range of 3 to 5: mean, 4.0	
Costs/year*		
Medical checks	€94	[12]
Drugs	Mild (donepezil): €187 Moderate (donepezil): €187 Severe (memantin): €579.4	[9]
Care	Mild €11,412 Moderate €22,470 Severe €25,867.2	[8]
Annual survival probability:	According to sex and age	[3]
	2.1 hazard ratio for people with AD applied	[4]
Treatment strategy:		[16]

Early detection	Diagnosed with mild AD, 100% treated.	
No early detection	Diagnosed with MMSE 19, 25% treated Consistent with the observations that only 25% of patients are diagnosed in Czechia.	[11]

Table 1: Overview of parameters used in Monte Carlo simulation

Our results are in line with the existing literature and suggest early detection and treatment of AD to be a cost-effective strategy. Similarly to Weimer's study, we present an example of a 70-year-old woman here (see **Figure 1**). Savings are realized in case she is diagnosed before the MMSE score falls under 16 (Lopez decline scheme) or 19 (Mean decline scheme) points, respectively. The earlier the diagnose comes and treatment begins the higher monetary savings. To better illustrate, the average lifetime costs in case of early diagnosis for a 70-year-old woman with MMSE=26 are €255,548 (299,950) according to Lopez (Mean) decline scheme. This amount can be divided according to its character into costs of medical checks €1,254 (1,251), costs of medication €4,257 (5,077) and costs of care €250,036 (293,622). She spent on average 4.9 (2.2) years living with mild, 2.8 (4.2) years living with moderate and 3.2 (4.6) years living with severe dementia. In comparison, the same person's expected lifetime costs without early detection are €273,995 (320,283); costs of medical checks are €137 (249), costs of medication €725 (1,404) and costs of care €273,133 (318,631). She spent 3.6 (1) years living with mild, 2.5 (2.7) living with moderate and 4.7 (7.2) years living with severe dementia. In both cases, she dies at age of 81 since treatment is not supposed to prolong lifetime, only to slow the disease progression [6]. Consequently, early detection saves in this particular case €18,488 (20,333), which is equivalent to approximately 500–550 thousand Czech Koruna (CZK). As could be seen from **Figure 2**, monetary savings following from early diagnosis (at MMSE equal to 26 points) decrease with the latter onset of the disease but stay positive even for very high ages.

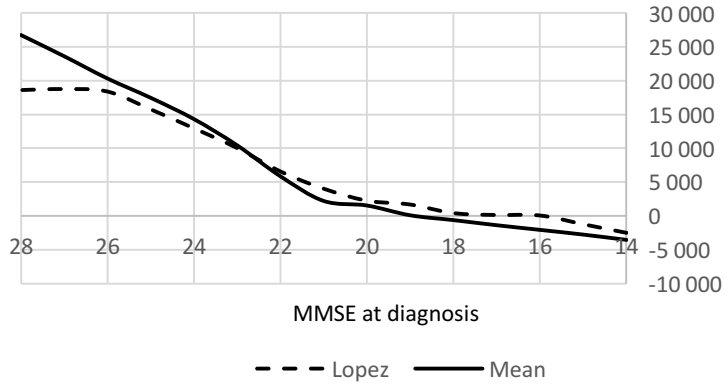


Figure 1 Monetary benefit (in EUR) of early diagnosis and treatment for a 70-year-old woman with AD

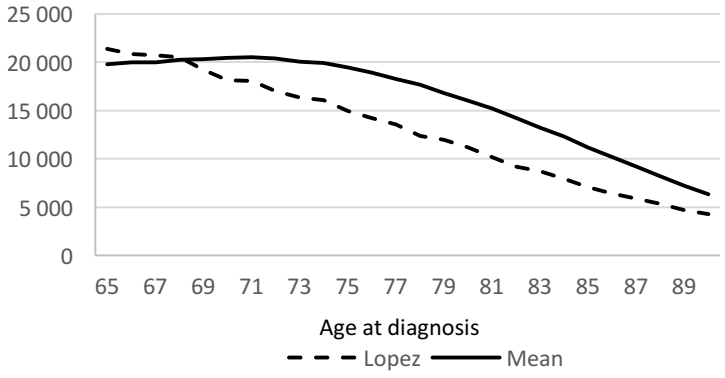


Figure 2 Monetary benefit (in EUR) of early diagnosis and treatment for a woman with MMSE of 26

4 Discussion

Models estimating cost-effectiveness inevitably rely on a number of assumptions. These assumptions should be reasonable representations of reality. As clear from section 2, there exists a trade-off between accuracy and difficult collection of data detailed enough to build a complex individual and events-based model. Here we discuss some of the assumptions and parameters used in Monte Carlo simulation for Czechia and how to ideally deal with them to build a more reliable model.

Let us voice three concerns. First, we use decline schemes that are not adjusted according to sex, age or severity of the disease. Moreover, these schemes are derived from a somewhat different treatment strategy for AD than we assume for Czechia (donepezil for mild and moderate AD, memantin for severe AD and fixed annual medical checks). To perform a reliable simulation, we need to estimate the probability distribution of MMSE scores (or another test used to measure severity of AD), develop a matrix of transition probabilities or to estimate probability that a patient stays in the same state. For that reason, we aim to collect and analyse patient-level data for Czechia to construct a context-specific decline scheme with adjustments for personal characteristics and comorbidities. This data will help us also to truly evaluate costs of medical checks and medication.

Second, we consider a constant death hazard ratio to be an unreliable assumption since severity of the disease is likely correlated with the death risk. In an ideal case, we aim to build Cox proportional hazard model and estimate death risk for AD patients in Czechia.

Third, the current MCS does not directly consider the institutionalization risk, but instead we use Czech estimates on costs of assistance in daily-activities for different intensities of the disease. Although introduction of the institutionalization risk into our model would likely even increase societal savings from early detection, this step is important nonetheless. While home care is mainly covered by personal funds together with a small share of state social benefits, costs of institutionalization, especially in mental hospitals, go usually to the expense of public health care budgets. Thus, the next step in our analysis – including introduction of the institutionalization risk – is to separate costs as well as possible savings according to who finances the care.

Although unlikely unbiased and certainly not detailed enough, our model based on Monte Carlo simulation gives some first idea regarding cost-effectiveness of Alzheimer's disease early detection. We aim to replace the most questionable assumptions by proper estimates based on a patients' sample that is large and detailed enough and to shift the simulation method towards discrete event simulation. By achieving this, we would be able not only to build a more reliable model giving accurate results but also to evaluate the severity of the bias inherent to the current model based on standardly used methodology and simplifying assumptions.

Acknowledgements

Supported by the grant No. F4/38/2017 of the Internal Grant Agency of the University of Economics, Prague.

References

[1] Banks, J, Carson II JS, Nelson, BL, and Nicol, DM.: *Discrete-Event System Simulation*, 4th edition. Prentice Hall, New Jersey, 2004.
[2] Caro JJ, Möller J, and Getsios D.: Discrete event simulation: the preferred technique for health economic evaluations? *Value Health* **13** (2010), 1056-1060.
[3] CZSO: Life tables for the Czech Repulic. Prague: Czech Statistical Office, 2016.
[4] Fitzpatrick AL, Kuller LH, Lopez OL, Kawas CH, and Jagust W.: Survival following dementia onset: Alzheimer's disease and vascular dementia. *J Neurol Sci* **229** (2005), 43-49.
[5] Getsios D, Blume S, Ishak KJ, Maclaine G, and Hernández L.: An economic evaluation of early assessment for Alzheimer's disease in the United Kingdom. *Alzheimer's & Dementia* **8** (2012), 22-30.
[6] Getsios D, Blume S, Ishak KJ, and Maclaine GD.: Cost effectiveness of donepezil in the treatment of mild to moderate Alzheimer's disease. *Pharmacoeconomics,* **28** (2010), 411-427.
[7] Green C, Shearer J, Ritchie CW, and Zajicek JP. Model-based economic evaluation in Alzheimer's disease: a review of the methods available to model Alzheimer's disease progression. *Value Health* **14** (2011), 621-630.
[8] Holmerová I, Hort J, Rusina R, Wimo A, and Šteffl M. Costs of dementia in the Czech Republic. *The European Journal of Health Economics* (2016), 1-8.
[9] Kruntoradova K, Mandelikova M, Mlcoch T, and Dolezal T.: Cost-Effectiveness Analysis of Ginkgo Biloba Extract (Egb761®-Tanakan®) for The Treatment of Dementia In The Czech Republic. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research* **18** (2015): A756-A756.

[10] Lopez OL, Becker JT, Saxton J, Sweet RA, Klunk W, and DeKosky ST.: Alteration of a clinically meaningful outcome in the natural history of Alzheimer's disease by cholinesterase inhibition. *J Am Geriatr Soc* **53** (2005), 83-87.
[11] Mátl O, and Mátlová M.: *Zpráva o stavu demence 2015*. Praha: Czech Alzheimer Society, 2015, p. 18.
[12] Mohelska H, Maresova P, Valis M, and Kuca K.: Alzheimer's disease and its treatment costs: case study in the Czech Republic. *Neuropsychiatr Dis Treat* **11** (2015), 2349-2354.
[13] Rubinstein RY and Kroese DP.: *Simulation and the Monte Carlo method*. John Wiley & Sons, Inc., New Jersey, 2008.
[14] Stahl JE.: Modelling methods for pharmacoeconomics and health technology assessment. *Pharmacoeconomics* **26** (2008), 131-148.
[15] Van Gestel A, Severens JL, Webers CA, Beckers HJ, Jansonius, NM, & Schouten, JS.: Modeling complex treatment strategies: construction and validation of a discrete event simulation model for glaucoma. *Value in Health*, **13.4** (2010), 358-367.
[16] Weimer DL, and Sager MA.: Early identification and treatment of Alzheimer's disease: social and fiscal outcomes. *Alzheimer's & Dementia* **5** (2009), 215-226.

Uncertainty in school examinations: Estimation of examinator's bias parameters

Ondřej Sokol[1], Michal Černý[2]

Abstract. This paper deals with biased observations coming from two sources. The task is to infer the distribution of the unobservable variables using two biased observations. The problem is illustrated by the example of an entrance examination. Each student takes a test which is anonymously evaluated by two examiners. Filled tests are randomly split up into several batches. Each batch is distributed to a pair of examiners who evaluate each test by points in a given range. Awarded points by examinators, which may be biased due to examinator attitude, represent observations. We assume that the observations are the sums of realizations of bias and true values. Under the assumption that true values, as well as the bias of each observer, have a normal distribution, our goal is to estimate the distribution parameters of the unobserved variables assuming reasonable assumptions in order to deliver identifiability.

Keywords: uncertainty, statistical inference, errors-in-variable.

JEL classification: C13
AMS classification: 62N02

1 Introduction

In this paper, we deal with a situation where true values are not directly observable, but what is available are somehow biased observations of that process from at least two sources. Under this setup, we want to infer the distribution of the hidden variables.

The problem is illustrated by the example of examination at school. Each student takes a test which is anonymously evaluated by two examiners. Filled tests are randomly split up into several batches. Each batch is distributed to (at least) a pair of examiners who evaluate each test by points in a given range. Awarded points by examinators represent observations. As some of the questions are open, awarded points are affected by examiner subjective feeling. To eliminate the subjective part of the evaluation process, we want to explore the bias of each examiner. We use only awarded points to estimate the properties of examiner's bias as the "true" results of tests are generally unknown.

Our illustration has roots in a real world problem which was recently discussed in the Czech Republic regarding high-school leaving exam as there were doubts about the examinators bias. The problem was revealed when teachers were assigned to mark test of students from different schools instead of its home school. The global results of high-school leaving exams decreased significantly at some schools.

Another problem is revelead by some "unexpected" behaviour of results distribution – for example the disproportionate number of students had exactly the minimal amount of points needed to pass. The problematic cases are usually the ones very close to score cutoff. Those are also the ones who are most likely to appeal againts the results in case of failure.

In this paper, we discuss one of the possible ways how to distribute tests to examinator in a way where is possible to estimate examinator bias as well as true – "unbiased" – distribution of exam's results. As a result, we could provide statistical data that could help in dealing with complaints regarding awarded points. We focus mainly on the estimating of variance as the mean can not be estimated without additional strong assumptions. Also, while we do not use real data, we try to set our parameters reasonable. The size of the sample is based on the real experiences of Czech high-school teachers, in particular teachers who are involved in the evaluation process of high-school leaving exams.

The article is organised as follows: In Section 2, we formally define the problem and its assumptions. In Section 3 we discuss the identifiability of mean under our setup and discuss the assumptions that allow us to estimate mean. In Section 4 we deal with estimating of variance using only the observable data. The paper concludes with a short summary in Section 5.

[1]University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, ondrej.sokol@vse.cz
[2]University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, cernym@vse.cz

2 Problem statement

Assume $N = \{1, 2, \ldots, n\}$ is the set of students and $M = \{1, 2, \ldots, m\}$ is the set of examiners. Examiners are assigned to pairs of two. Let x_i for $i \in N$ is the test results of i-th student in points which is not observable. What we observe are awarded points a_{ij} and a_{ik} by at least two different examiners $j \in M$ and $k \in M$ ($j \neq k$) who evaluated the test. These examiners j and k who were apriori assigned to pair are not unbiased. Some of them are more likely to give more points, some have opposite attitude. The biases are not known, but we observe how they evaluate tests assigned to them.

Note: We restrict ourselves to the case where only one pair of examinators marks the test. If there would be more examinator assigned to one test, we could make more precise estimation by combining estimates from all possible pairs of examinators; however, that situation would be too expensive in practice.

Assumption 1. Let

$$x_i \sim N(\mu, \sigma^2), \quad i = 1, \ldots, n, \tag{1}$$

where x_i are independent for $i = 1, \ldots, n$. This random variable is unobservable. What is observable for every i are at least two results of following random processes:

$$a_{ij} = x_i + \alpha_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m \tag{2}$$

where $\alpha_{ij} \sim N(\nu_j, \pi_j^2)$ for each $i = 1, \ldots, n$ and $j = 1, \ldots, m$ and x_i, α_{ij} for each $i = 1, \ldots, n$ and $j = 1, \ldots, m$ are independent.

In this case, we assume for simplicity that there are exactly two observable variables a_{ij} for each $i = 1, \ldots, n$; however, there can be an arbitrary number of observable variables of each i if they follow the same generation process (see 2). In that case, our approach can be analogically extended. In our example that would mean that there can be any size of group such that is granted that every test is evaluated by at least two examiners.

Our goal is to estimate parameters μ and $\nu_1, \nu_2, \ldots, \nu_m$ and σ^2 and $\pi_1^2, \pi_2^2, \ldots, \pi_m^2$ using only a collection of observable values of a_{ij} which are availabe for certain combinations of $[i; j] \in [\{1, \ldots, n\}; \{1, \ldots, m\}]$. To put it in context with our example, we want to find out the expected value and variance of true test results and the expected value and variance of bias of each examiner.

Our problem is illustrated in Figure 1 where we plotted histogram of sample of 200 observations and theoretical distributions of both observable and one unobservable variables. Our goal is to estimate parameters of unobservable distributions using only observable samples. It is important to note that examiners are divided into pairs (or larger

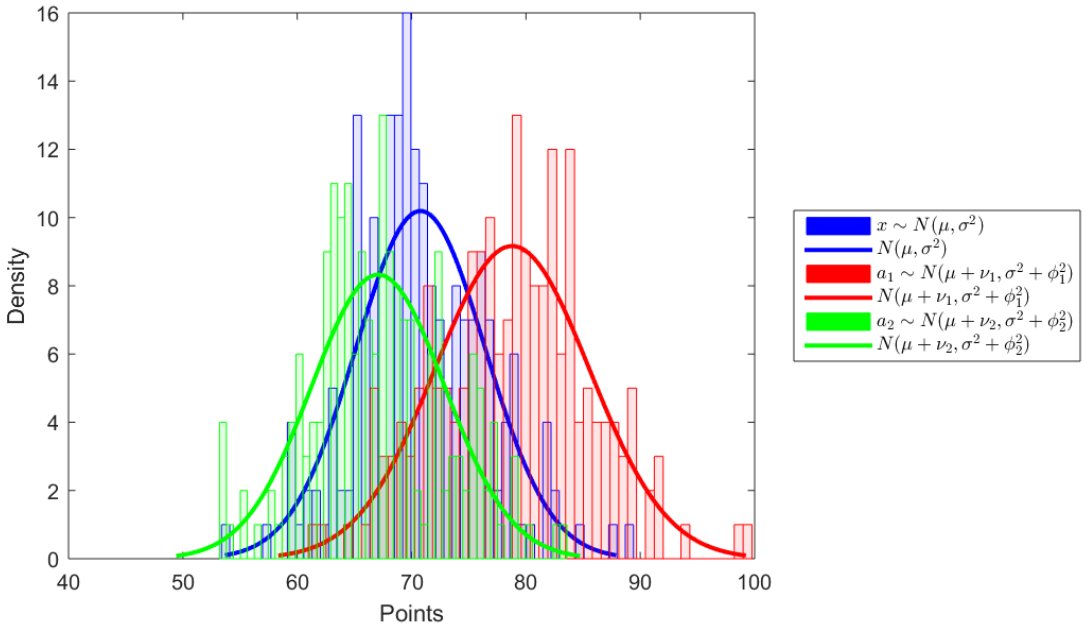


Figure 1 Histogram of sample distribution (200 observations) and random variable distribution of both observable (red and green) and unobservable (blue) variables.

group), and whole batches of tests are assigned to those pairs (or groups) to reduce sampling error. We focus only on expected value and variance of examinators' biases and "true" results.

Our problem is similar to random effect models (see for example [6]) or to error-in-variables models (see for example [2] and [3]). However, to the contrary of general error-in-variables or random effect models, we have a very specific setup that allows us to estimate parameters of generally unobservable random variables distributions using only the observable ones.

3 Estimating mean

3.1 Point estimation

The first goal is to estimate μ and $\nu_1, \nu_2, \ldots, \nu_m$ from the observable data. It is obvious that

$$\begin{aligned} \mathsf{E}(a_{i1}) &= \mu + \nu_1 \\ \mathsf{E}(a_{i2}) &= \mu + \nu_2 \\ &\vdots \\ \mathsf{E}(a_{im}) &= \mu + \nu_m \end{aligned} \tag{3}$$

Let

$$A_j = \frac{1}{n_j} \sum_{i \in N_j} a_{ij}, \tag{4}$$

where N_j is a subset of $\{1, 2, \ldots, n\}$ indicating which tests were evaluated by j-th examinator and $n_j = |N_j|$.

Given that there is a sufficient number of observable variables a_{ij} for every j, we have following equation system:

$$\begin{aligned} A_1 &= \widehat{\mu} + \widehat{\nu_1} \\ A_2 &= \widehat{\mu} + \widehat{\nu_2} \\ &\vdots \\ A_m &= \widehat{\mu} + \widehat{\nu_m} \end{aligned} \tag{5}$$

The equation system 5 consist of m equations and $m + 1$ variables. It is not possible to estimate μ and $\nu_1, \nu_2, \ldots, \nu_m$ from the observable data as the problem is underdetermined. Therefore we need to add more assumptions. There are two reasonable assumptions which we explore in the following paragraphs.

3.2 Assumptions to deliver identification

Firstly, we assume that the mean value of bias of examiners as a group is some constant, e. g. examiners as a group are not biased and mean of their expected value is zero. Formally we can write it in the following way.

Assumption 2. Let

$$\frac{1}{n} \sum_i \nu_i = c \tag{6}$$

where c is a known constant.

To put it in the context of our example, this Assumption 2 presumes that the examiners are unbiased as a group.

Assumption 3. Let the following equation is valid:

$$\nu_1 = c \tag{7}$$

where c is given constant.

For $c = 0$ this Assumption 3 would mean that we can estimate bias of at least one examiner – for example one of the examiners is the author of the test who defined the evaluation process of the exam. Hence his bias is zero.

Adding the first Assumption 2 in the form of

$$\frac{1}{n} \sum_j \widehat{\nu_j} = c \tag{8}$$

or the second Assumption 3 in the form of

$$\widehat{\nu_1} = c \tag{9}$$

to the equation system 5 allows us to estimate μ and $\nu_1, \nu_2, \ldots, \nu_m$ by solving resulting equation system.

3.3 Partial identification approach

We can also deal with the problem in different manner without using strong assumptions 2 and 3. This idea is based on partial identification (see [4] or [1]). In this case we do not assume any exact formula of examinator's bias but we assume bounds of average bias $\underline{\nu}$ and $\overline{\nu}$ such that

$$\underline{\nu} \leq \frac{1}{n} \sum_i \nu_j \leq \overline{\nu} \tag{10}$$

where $\underline{\nu}$ and $\overline{\nu}$ is the lowest (highest) average bias we assume. In that case, without further assumption, our goal is to find all feasible solutions for equation system 5 subject to

$$\underline{\nu} \leq \frac{1}{n} \sum_j \widehat{\nu_j} \leq \overline{\nu}. \tag{11}$$

As a result a set of feasible solutions is found. Using feasible set we can compute the intervals $[\underline{\widehat{\nu}_j}, \widehat{\overline{\nu}_j}]$ of each bias $\widehat{\nu_j}$ and interval $[\underline{\widehat{\mu}}, \widehat{\overline{\mu}}]$ of $\widehat{\mu}$ such that

$$\underline{\widehat{\nu}_j} \leq \widehat{\nu_j} \leq \widehat{\overline{\nu}_j}, \quad j = 1, \ldots, m \tag{12}$$

$$\underline{\widehat{\mu}} \leq \widehat{\mu} \leq \widehat{\overline{\mu}}. \tag{13}$$

The resulting intervals $[\underline{\widehat{\nu}_j}, \widehat{\overline{\nu}_j}]$ for all $j \in M$ and $[\underline{\widehat{\mu}}, \widehat{\overline{\mu}}]$ can be used to infer about true test results of individual students.

4 Estimating variance

4.1 Two-phase approach

In this section, we focus on the estimation of the variance of unobservable variables. We present an approach where we do not use expected values which we were dealing with in a previous section. Thus we do not require to adding assumptions 2 or 3 to estimate variance. We propose a two-phase approach. In the first phase we estimate the parameters independently for each batch and in the second phase we use all of the observable information. We use only basic variance formulas, see for example [5].

Assume we have two examiners j and k and one batch of tests, which is distributed among them. Our first goal is to estimate variance parameters $\sigma^2, \pi_j^2, \pi_k^2$ from each batch of tests independently. To estimate σ^2, we can use sample covariance of observable values – awarded points by examinators.

We use the covariance of random variables a_{ij}, a_{ik} along with assumptions that α_{ij} and α_{ik} for $i \in N$, $j, k \in M$, $j \neq k$ are independent and formula for the expected values of squares of the normal variable

$$\mathsf{E}(x^2) = \mu^2 + \sigma^2 \tag{14}$$

Then the covariance of the observable values a_{ij}, a_{ik} is following

$$\begin{aligned}
\mathsf{cov}(a_{ij}, a_{ik}) &= \mathsf{E}(a_{ij}a_{ik}) - \mathsf{E}(a_{ij})\mathsf{E}(a_{ik}) \\
&= \mathsf{E}\left[(x_i + \alpha_{ij})(x_i + \alpha_{ik})\right] - \mathsf{E}\left[x_i + \alpha_{ij}\right]\mathsf{E}\left[x_i + \alpha_{ik}\right] \\
&= \mathsf{E}\left[x_i x_i + x_i\alpha_{ij} + x_i\alpha_{ik} + \alpha_{ij}\alpha_{ik}\right] - (\mu + \nu_j)(\mu + \nu_k) \\
&= (\mu^2 + \sigma^2 + \mu\nu_j + \mu\nu_k + \nu_j\nu_k) - (\mu^2 + \mu\nu_j + \mu\nu_k + \nu_j\nu_k) \\
&= \sigma^2.
\end{aligned} \tag{15}$$

We use the implication from the last paragraph to estimation of σ^2

$$\widehat{\sigma^2_{(j,k)}} \coloneqq \widehat{\mathsf{cov}}(a_{ij}, a_{ik}) \tag{16}$$

where $\widehat{\sigma^2_{(j,k)}}$ is the estimated value of σ^2 using only observable data from pair of examinators $[j, k]$ and $\widehat{\mathsf{cov}}(a_{ij}, a_{ik})$ is sample covariance from observable data.

In the first phase, we independently estimate parameters $\sigma^2, \pi_j^2, \pi_k^2$ for each pair of examiners $[j,k]$. Therefore we get a collection of estimates $\widehat{\sigma^2_{(j,k)}}, \widehat{\pi_j^2}, \widehat{\pi_k^2}$ for every pair $[j,k]$ which evaluated at least one test together. While the estimates are unbiased, the sampling error for estimating variance is usually quite high. For that reason, we propose the second phase to reduce the sampling error at least for estimation of σ^2. Therefore in the second phase, we use estimates from the previous paragraph to reduce the sampling error of σ^2.

Using the assumption that μ is the same in all the samples, we can estimate the σ^2 using information from all observations and thus reducing the sample error.

$$\widehat{\sigma^2} := \mathsf{E}\widehat{\sigma^2_{(j,k)}} = \frac{1}{\sum_{[j,k]\in M} n_{(j,k)}} \sum_{[j,k]\in M} n_{(j,k)} \widehat{\sigma^2_{(j,k)}} \tag{17}$$

where $n_{(j,k)}$ is the number of tests assigned to pair j and k examinators. As a result, we have used information from every evaluated test to obtain the estimate of σ^2. As the only operation we did was sum variances of independent variables we can also compute confidence interval for σ^2 using standard procedures

$$\frac{(n-1)\widehat{\sigma^2}}{\chi^{2\,(1-\alpha/2)}_{(n-1)}} \leq \sigma^2 \leq \frac{(n-1)\widehat{\sigma^2}}{\chi^{2\,(\alpha/2)}_{(n-1)}} \tag{18}$$

where $\widehat{\sigma^2}$ is in fact the sample variance and α stands for the desired significance level.

With more accurate estimate $\widehat{\sigma^2}$ we can recalculate estimation of bias of each examinator j:

$$\widehat{\pi_j^2} := \widehat{\mathsf{var}_i a_{ij}} - \widehat{\sigma^2}. \tag{19}$$

However, due a insufficient number of observations in practical applications, sampling error can be large enough resulting in $\widehat{\pi_j^2}$ being negative, which is not a valid result for estimation of variance.

4.2 Simulation

We tested the method using random data. In Figure 2 we show a simulation study of the estimator with 20 examinators (10 pairs of 2) and 200 tests assigned to each pair.

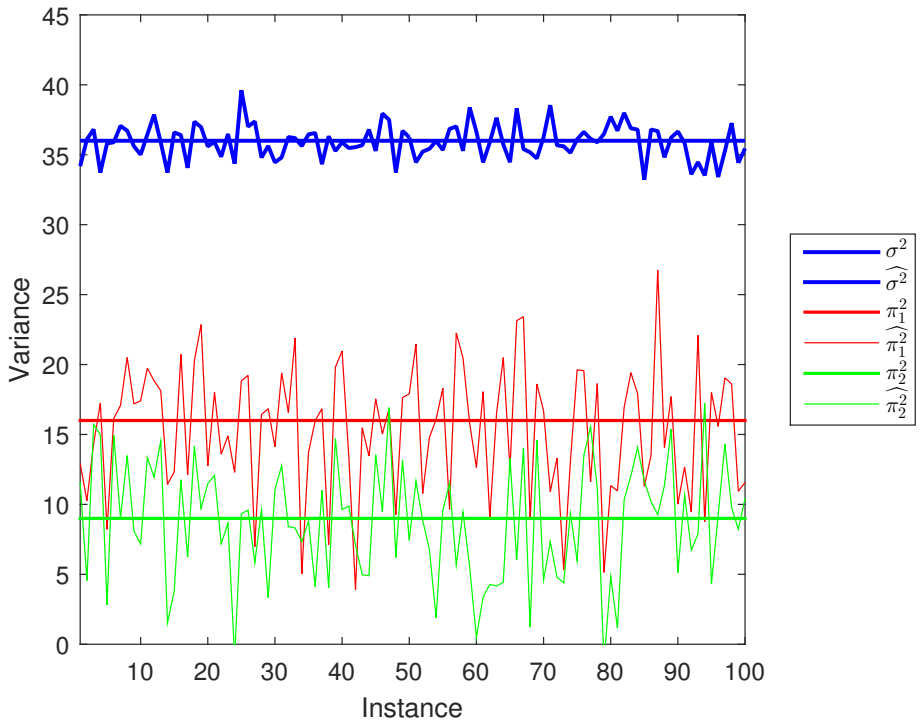


Figure 2 Comparison of estimated and true variance in randomly generated instances with $m = 20$ (10 pairs) and $n_{jk} = 200$ for all 10 pairs $[j,k]$. For clarity variance of only pair $[1,2]$ is shown.

The standard error of examinators was set randomly to integer between 1 and 10. We run the simulation 100 times. We can see that the estimates of $\widehat{\sigma^2}$ are quite accurate. On the other hand, estimates of bias of individual examinators are more affected by sampling error. The reason is that estimation of variance needs a relatively large amount of data.

5 Conclusion

In this paper, we dealt with the situation where the "true" values are unobservable, but at least two biased observations are available. Under the assumption that all variables have the normal distribution and observable biased variables are the sum of "true" value and bias realization, we presented an approach to estimate the expected value and variance of both "true" values distribution and bias distribution under the particular setup inspired by evaluation of tests at school (such as entrance examination or high-school leaving exam).

We showed that the estimation of the mean cannot be identified without additional strong assumption. Various possible reasonable assumptions were discussed, and non-point-identified approach inspired by partial identification was proposed. On the contrary of the mean estimation, the variance parameters can be estimated from observable variables given that large enough sample is available. We showed the dependence of variance on other directly computable statistics which can be used for unbiased variance estimation.

Acknowledgements

The work of Ondřej Sokol was supported by the grant No. F4/63/2016 of the Internal Grant Agency of the University of Economics, Prague. The work of Michal Černý was supported by the Czech Science Foundation under Grant 16-00408S.

References

[1] Chernozhukov, V., Hong, H., and Tamer, E.: Estimation and Confidence Regions for Parameter Sets in Econometric Models. *Econometrica* **75** (2007), 1243–1284.

[2] Griliches, Z.: Errors in Variables and Other Unobservables. *Econometrica* **42** (1974), 971.

[3] Griliches, Z., and Hausman, J. A.: Errors in variables in panel data. *Journal of Econometrics* **31** (1986), 93–118.

[4] Manski, C. F.: *Partial Identification of Probability Distributions*. Springer Science & Business Media, 2003.

[5] Rényi, A.: *Probability Theory*. Dover ed edition. Dover Publications, Mineola, N.Y, 2007. OCLC: ocm77708371.

[6] Searle, S. R., Casella, G., and McCulloch, C. E.: *Variance Components.* John Wiley & Sons, Inc., Hoboken, 2009. OCLC: 781254318.

Future performance of mean-risk optimised portfolios: an empirical study of exchange traded funds

Tomas Spousta[1], Adam Borovicka [2]

Abstract. The paper investigates whether mean-risk optimal portfolios achieve better out of the sample performance than naive strategies (e.g. 1/N diversification). Results are tested on both simulated and real data. Real dataset consists of adjusted daily returns of 26 exchange-traded funds (ETFs) in years 2005-2016. Various sectors (e.g. financial, technology, industrial, health care, utilities), countries (e.g. US, UK, Japan, Germany, Australia, South Korea), and asset classes (stocks, bonds, commodities, real estate) are represented in selected ETFs. Weights of individual ETFs are estimated from historical data with respect to realized return and risk. Variance and Conditional Value at Risk (CVaR) are assumed as risk measures and mean-risk optimal portfolios are then compared to 1/N strategy on out of the sample dataset where its performance is discussed in detail.
Both algorithms outperformed 1/N strategy on risk-adjusted basis on simulated dataset of stable multivariate-normally distributed returns. However, performance of tested models were not superior to 1/N diversification on the real dataset. In both models, lower required return led to better outsample performance measured on a risk-adjusted basis.

Keywords: Portfolio optimisation, mean-variance, mean-CVaR, efficient portfolio frontier, 1/N diversification.

JEL classification: C44
AMS classification: 90C90

1 Introduction

The problem of wise diversification of investments attracts attention of many researchers from academia and practise. The idea is to determine a set of portfolios that will carry the lowest risk among all feasible portfolios at a given level of expected return. This paper compares performance of two well-established mean-risk models (mean-variance and mean-CVaR) in portfolio selection of exchange-traded funds.

Mean-variance model, originally proposed by Markowitz in 1952 [2], has become state-of-the-art approach in portfolio theory. However, some unwanted properties of standard deviation (e.g. lack of monotony and thus lack of coherence [1], or penalisation of positive deviations from the expected return) led researchers to search for new risk measures. One class of risk measures that has become very popular in recent years is based on quantiles. Conditional value at risk (CVaR, [5]) is probably the most popular representative of quantile-based risk measures. Main advantages of CVaR are straightfoward interpretation (average loss that occurs in α-% worst cases), coherence as defined in [1], and robustness to extreme values of returns that are typically present in financial data. Both models mentioned above are compared with naive benchmark portfolio composed of equal investment in all securities (1/N strategy).

Mean-variance and mean-CVaR optimisation problems are defined in the following section. In the application part, models are tested on simulated multivariate normal returns with stable parameters and on real historical prices of 26 ETFs in years 2005–2016.

2 Models

Both portfolio selection models are formulated in this section. Mean-variance model formulation is adopted from [2] whereas mean-CVaR model is taken from [5] where both models are described in more detail including their assumptions, derivation, and interpretation of results.

2.1 Mean-variance model

Mean-variance optimal portfolio is a solution of following optimization problem:

[1]University of Economics, Department of Econometrics, W. Churchill Sq. 1938/4, 130 67 Praha 3, Czech Republic, xspot00@vse.cz
[2]University of Economics, Department of Econometrics, W. Churchill Sq. 1938/4, 130 67 Praha 3, Czech Republic, adam.borovicka@vse.cz

$$\begin{aligned} & \underset{\mathbf{x}\in X}{\text{minimize}}\{\mathbf{x}^T\mathbf{V}\mathbf{x}\} \\ \text{s. t.} \quad & \mathbf{x}^T\mathbf{m} \geqslant R, \\ & \sum_{i=1}^{n} x_i = 1, \\ & x_i \geqslant 0,\ i = 1, 2, \ldots, n, \end{aligned} \tag{1}$$

where x_i represents share of the *i*-th asset in portfolio, **m** is a vector of expected returns, **V** is covariance matrix of returns, and *R* is a constant representing required expected return. Solving the problem for various values of *R* results in different portfolios along the efficient potrfolio frontier.

2.2 Mean-CVaR model

Mean-CVaR optimal portfolio is a solution of following optimization problem:

$$\underset{(\mathbf{x},\beta)\in X\times\mathbb{R}}{\text{minimize}}\{\beta + \frac{1}{q\alpha}\sum_{k=1}^{q}[f(\mathbf{x},\mathbf{y}_k) - \beta]^+\} \tag{2}$$

$$\begin{aligned} \text{s. t.} \quad & \mathbf{x}^T\mathbf{m} \geqslant R, \\ & \sum_{i=1}^{n} x_i = 1, \\ & x_i \geqslant 0,\ i = 1, 2, \ldots, n, \end{aligned}$$

where x_i represents share of the *i*-th asset in portfolio, **m** is a vector of expected returns, and *R* is a constant representing required expected return. Solving the problem for various values of *R* results in different portfolios along the efficient potrfolio frontier. $\mathbf{y}_k$ is a vector of returns for *k*-th period, *q* is a total number of observed periods, β is a parameter that represents loss treshold, and $[t]^+ = t$ for $t > 0$ otherwise $[t]^+ = 0$. α denotes a quantile of loss of the CVaR (in the application part $\alpha = 0.05$ which means the average loss is computed from 5% of the worst days in the observed time period).

3 Application to exchange traded funds

Methods proposed in section 2 are applied to real dataset in the following part of the paper. Source of the data is Yahoo! [8] and the analysis is done in R [4] and MATLAB [3]. The dataset consists of adjusted daily returns of 26 ETFs in years 2005–2016. Models proposed in part 2 were compared to each other and to naive diversification (equal weigt of all securities) on outsample data. Various sectors (financial, technology, industrial, health care, utilities,...), countries (US, UK, Japan, Germany, Australia, South Korea,...), and asset classes (stocks, bonds, commodities, real estate) are represented in selected ETFs listed in table 1.

Ticker	Name of the ETF	Category	Ticker	Name of the ETF	Category
XLB	Materials Select Sector SPDR	Equity (USA)	EWD	iShares MSCI Sweden ETF	Equity (world)
XLE	Energy Select Sector SPDR	Equity (USA)	EWG	iShares MSCI Germany ETF	Equity (world)
XLF	Financial Select Sector SPDR	Equity (USA)	EWH	iShares MSCI Hong Kong ETF	Equity (world)
XLI	Industrial Select Sector SPDR	Equity (USA)	EWJ	iShares MSCI Japan ETF	Equity (world)
XLK	Technology Select Sector SPDR	Equity (USA)	EWT	iShares MSCI Taiwan ETF	Equity (world)
XLP	Consumer Staples Select Sector SPDR	Equity (USA)	EWU	iShares MSCI United Kingdom ETF	Equity (world)
XLU	Utilities Select Sector SPDR	Equity (USA)	EWW	iShares MSCI Mexico Capped ETF	Equity (world)
XLV	Health Care Select Sector SPDR	Equity (USA)	EWY	iShares MSCI South Korea Capped ETF	Equity (world)
XLY	Consumer Discretionary Select Sector SPDR	Equity (USA)	EWZ	iShares MSCI Brazil Capped	Equity (world)
IBB	Nasdaq Biotechnology	Equity (USA)	LQD	iShares iBoxx $ Investment Grade Corporate Bond	Bonds
IJT	iShares S&P Small-Cap 600 Growth ETF	Equity (USA)	IEF	iShares 7-10 Year Treasury Bond	Bonds
EWA	iShares MSCI Australia ETF	Equity (world)	GLD	SPDR Gold Shares	Gold
EWC	iShares MSCI Canada ETF	Equity (world)	IYR	iShares US Real Estate	Real Estate

Table 1 List of selected ETFs

3.1 Simulated returns

In the first step, performance of mean-variance and mean-CVaR models is tested on simulated data. Average return of each security and corresponding covariance matrix were computed, then eigenvalue decomposition method [7] was employed to generate joint-normally distributed returns reflecting the observed returns and correlation

structure. Parameters of the joint-normal distribution were assumed constant over time. Dataset of 20 000 daily returns was generated with this procedure and split into 50 subsets of 400 observations. Weights of individual securities were estimated separately on each subset and model performance was tested on following (outsample) subset. Both models are compared in terms of average outsample return and Sharpe ratio [6] with zero risk free rate.

Figure 1 shows the average characteristics over all subsets. The figure shows that outsample performance of mean-variance and mean-CVaR models is very similar and both models achieved higher Sharpe ratio than benchmark 1/N strategy. Both models achieved the highest outsample risk-adjusted return for relatively lower values of required return R.

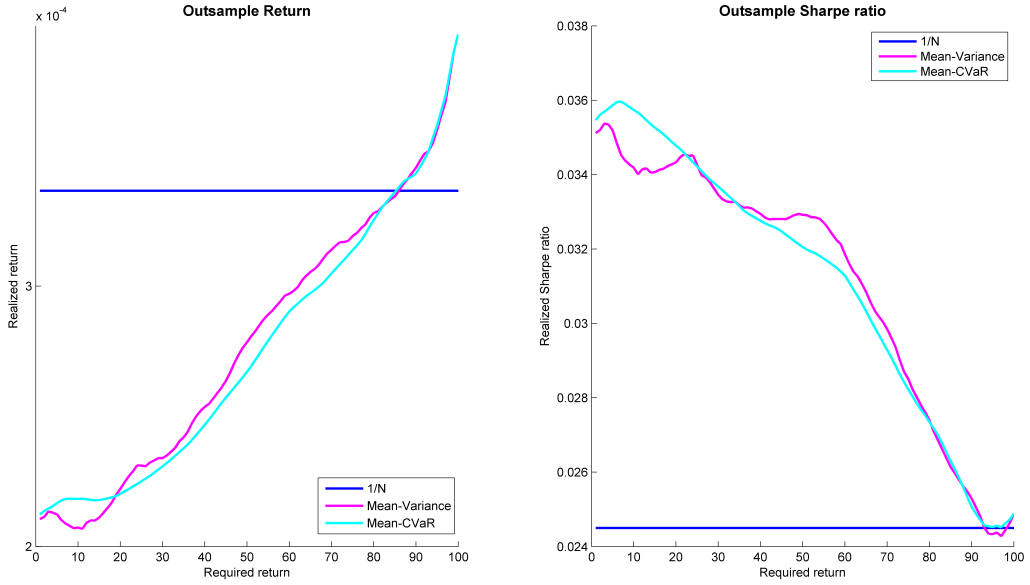


Figure 1 Results Summary (simulated returns)

Figures 2 and 3 show detailed results for all subsets separately. Color represents observed outsample statistic for given level of required expected return (horizontal axis), and for particular subset (vertical axis). Figures reveal that outsample performance of mean-variance and mean-CVaR models is more stable with portfolios of relatively lower levels of required expected return.

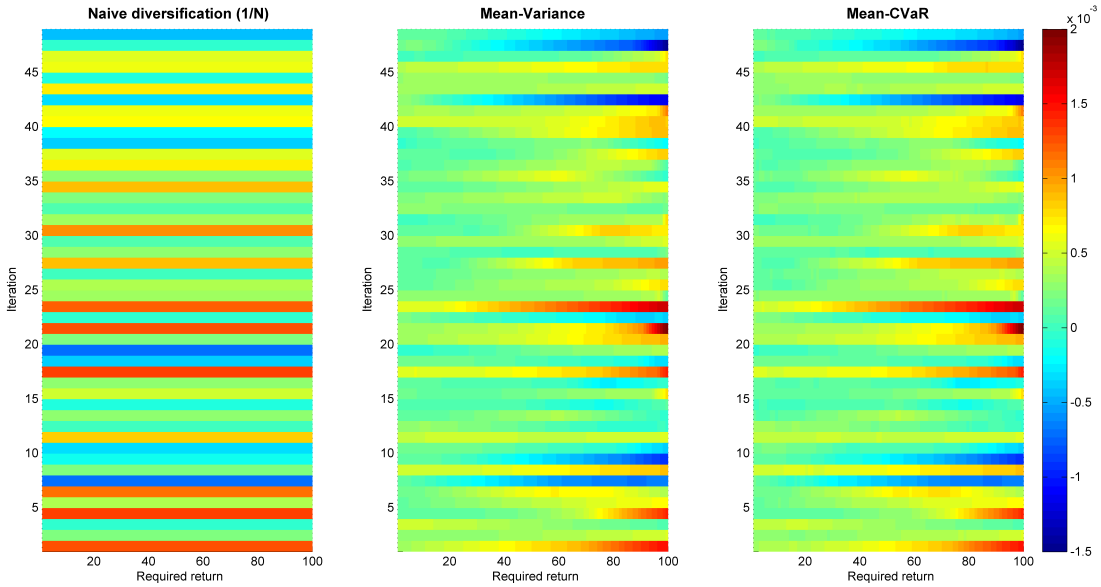


Figure 2 Outsample returns (simulated values)

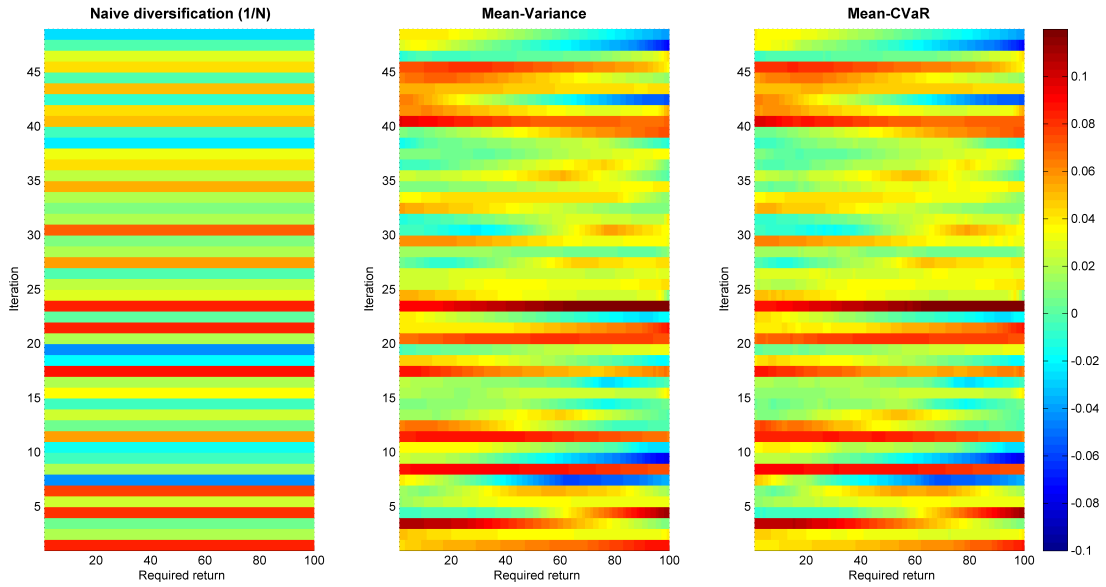


Figure 3 Outsample Sharpe ratio (simulated values)

The analysis conducted on simulated dataset shows that implementation of portfolio optimization models is worth of consideration as both models were able to outperform naive 1/N benchmark on risk-adjusted basis. However, one should note that simulated returns used in this part were generated from joint-normal distribution with stable parameters which is very strong assumption in confrontation with reality observed on financial markets.

3.2 Observed returns

In this section, models are applied to real daily returns of 26 selected ETFs. Whole time period was split into 14 blocks of 200 observations and portfolio weights were estimated on each block separately. Finally, performance of assessed models was measured on subsequent blocks. As in the application to simulated returns, outsample return and outsample risk-adjusted return are evaluated.

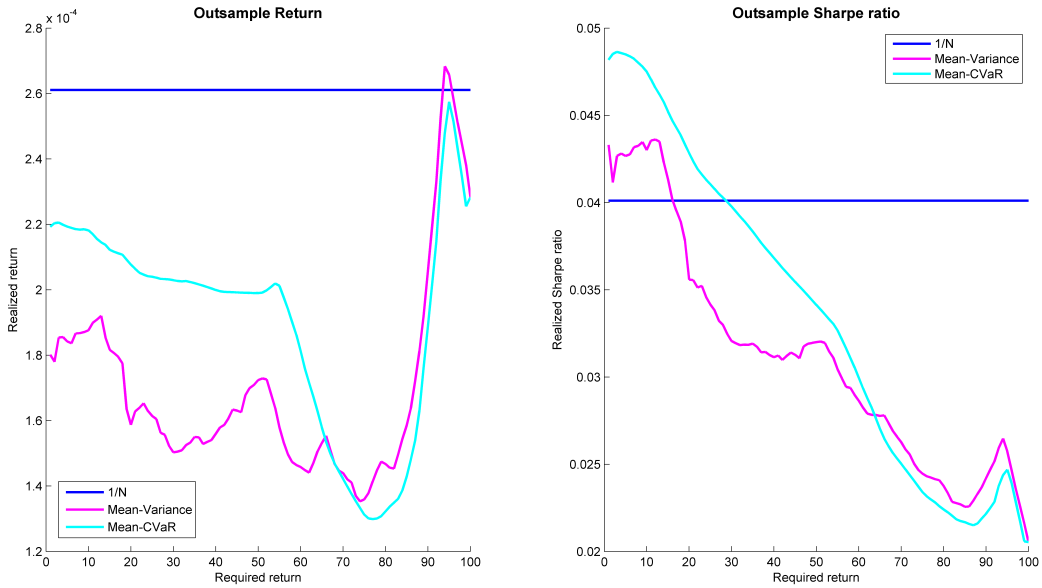


Figure 4 Results Summary (observed returns)

Figure 4 shows the average outsample characteristics. Results are very similar for both tested models, nevertheless mean-CVaR model slightly outperformed mean-variance approach in case of relatively lower required

expected return. Models achieved higher risk-adjusted return with portfolios of lower level of required expected return R, which is in line with results obtained on simulated data. However, both models were not able to deliver better results than naive 1/N strategy in longterm horizon.

Figures 5 and 6 show detailed results for all subsets separately. Color represents observed outsample statistic for given level of required expected return (horizontal axis), and for particular subset (vertical axis). Figures reveal that outsample return of mean-variance and mean-CVaR models is more stable with portfolios of relatively lower levels of required expected return and these portfolios on average achieve higher risk-adjusted returns on given dataset. Results of both models were more volatile with real dataset than with simulated values.

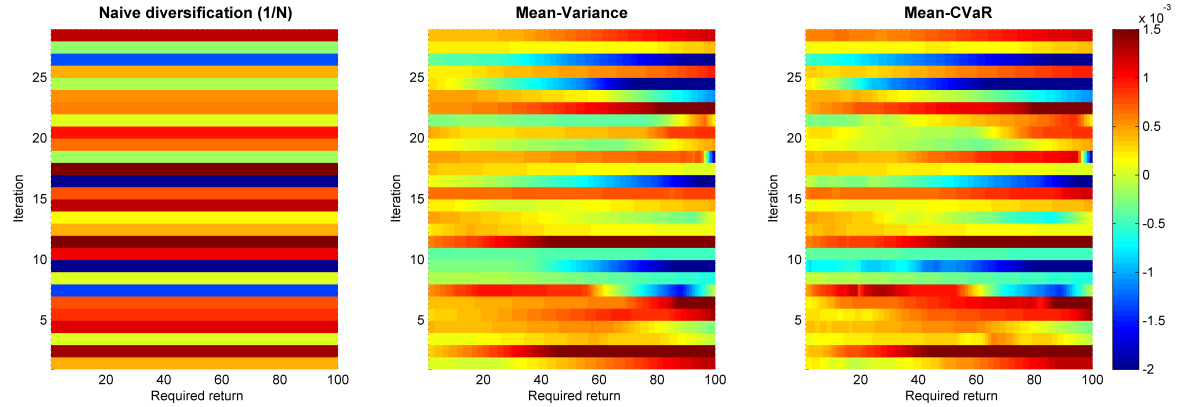


Figure 5 Outsample returns (observed values)

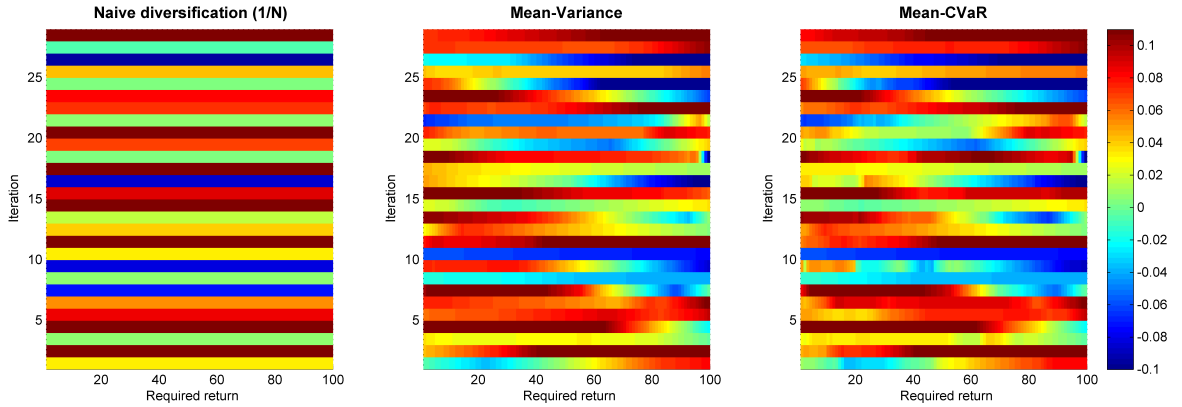


Figure 6 Outsample Sharpe ratio (observed values)

The analysis of ETF returns data does not support the conclusion from previous section as mean-risk models does not perform better than naive strategy on this particular dataset. The reason why would be an interesting topic for further research. One of the possible reasons could be the time instability of probability distribution of returns. On the other hand, the analysis suggests that investor aiming at maximization of risk-adjusted return should opt for relatively lower level of required return R resulting in portfolio weights close to minimal-risk portfolio.

4 Conclusion

Main goal of the paper was to test whether mean-variance and mean-CVaR optimal portfolios achieve better out of the sample performance than naive strategies (e.g. 1/N diversification). Results were tested on both simulated and real data.

An interesting finding is models achieved higher risk-adjusted return on portfolios of the low-risk part of the efficient frontier on both simulated and real data. On given real dataset mean-CVaR tends to slightly better results than mean-variance whereas no major difference in performance of both models was identified on simulated data.

Evidence from exchange traded funds shows both mean-variance and mean-CVaR models outperform 1/N strategy in imaginary world where asset returns follow multinomial normal distribution with stable expected return,

variance, and correlation structure. However, mean-risk models were not able to achieve higher risk-adjusted return on real dataset. Reasons of failure in real dataset would be an interesting topic for further research. One possible hypothesis is time instability of expected return, (co)variance, and higher moments of asset returns.

Acknowledgements

Supported by the grant No. IGA 57/2017 of the Internal Grant Agency of the University of Economics, Prague.

References

[1] Artzner, P., Delbaen, F., Eber, J. and Heath, D.: Coherent Measures of Risk. *Mathematical Finance* **9** (1999), 203–228.

[2] Markowitz, H.: Portfolio Selection. *The Journal of Finance* **7** (1952), 77–91.

[3] MATLAB Release 2013b, The MathWorks, Inc., Natick, Massachusetts, United States.

[4] R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016. Retrieved April 30, 2017, from https://www.R-project.org/.

[5] Rockafellar, T., Uryasev, S.: Optimization of Conditional Value-at-Risk. *The Journal of Risk* **2** (2000), 21–42.

[6] Sharpe, W. F.: The Sharpe Ratio. *The Journal of Portfolio Management*, **21**, 1994, 49-58.

[7] Venables, W. N., Ripley, B. D.: *Modern Applied Statistics with S*. Fourth Edition, Springer, New York, 2002.

[8] Yahoo Finance, Retrieved April 30, 2017, from http://finance.yahoo.com.

On the impact of correlation between variables on the accuracy of calibration estimators

Tomasz Stachurski[1], Tomasz Żądło[2]

Abstract. We consider three types of calibration estimators: Generalized Regression Estimator (GREG) and Modified Generalized Regression Estimator (MGREG) of subpopulation totals and GREG of the population total. The main purpose of this paper is to verify in the simulation study which levels of the correlation dependence between auxiliary variables and the variable of interest as well as only between auxiliary variables give both gains in accuracy of calibration estimators and low values of biases of estimators of their variances. We also analyze an impact of the number of auxiliary variables on biases and estimation accuracy.

Keywords: survey sampling, calibration estimators, auxiliary variables

JEL Classification: C83
AMS Classification: 62D05

1 Introduction

In economic research there is often a need to estimate characteristics such as mean or total value in population. Nowadays, decision-makers are not only interested in population characteristics, but also in a variety of subpopulations (domains) parameters. There are a few approaches used in survey sampling. In the design-based approach values of the study variable are treated as fixed and properties of the estimators are studied with respect to the sampling distribution. In the model-based approach values of the variable of interest are treated as realizations of random variables and the assumptions about their distribution are called the superpopulation model. In this approach the inference is usually conditional on the sample and properties of predictors are analyzed with respect to the distribution of the random variables. These approaches are presented widely e.g. in [6]. They can also be studied together (see e.g. [4]). In the paper we study properties of calibration estimators known from the model-assisted approach. Calibration estimators usually have good both design-based (e.g. design-consistency) and model-based (e.g. model-unbiasedness) properties, but their accuracy is measured in regard of the sampling design, so this approach is more similar to the design-based approach rather than the model-based approach. The main purpose of this paper is to analyze in the simulation studies properties of some calibration estimators and estimators of their variances to explore the influence of supporting estimation by auxiliary variables. Results presented in the paper may be interesting directly for statistical offices, market research and opinion polling companies and indirectly for decision-makers and all data users.

2 Calibration estimators of population total

From the population Ω of size N we draw at random a sample s of size n using sampling design denoted by $P(s)$ with first and second order inclusion probabilities denoted by π_i and π_{ij}, respectively. The generalized regression estimator of the population total, which belongs to the class of calibration estimators, is given by (see [1]; [9] p. 232, [6] p. 13):

$$\hat{\theta}^{GREG} = \sum_{i \in s} w_{si} y_i \,, \tag{1}$$

where

$$w_{si} = d_i + (\sum_{i \in \Omega} \mathbf{x}_i - \sum_{i \in s} d_i \mathbf{x}_i)^T \left(\sum_{i \in s} d_i q_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \mathbf{x}_i d_i q_i \,, \tag{2}$$

$\mathbf{x}_i$ is $p \times 1$ vector of auxiliary variables for ith population element, design weights are denoted by $d_i = \pi_i^{-1}$, and

1 University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, 50, 1 Maja Street, Katowice, Poland, e-mail: tstachurski93@gmail.com
2 University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, 50, 1 Maja Street, Katowice, Poland, e-mail: tomasz.zadlo@ue.katowice.pl

where for more generality additional known weights q_i can be included.

The formula of weights (2) can be obtained as a solution of the following conditional minimization problem ([1]):

$$\begin{cases} f_s(w_{si}, d_i, q_i) \to \min \\ \sum_{i \in s} w_{si}\mathbf{x}_i = \sum_{i \in \Omega} \mathbf{x}_i \end{cases}, \tag{3}$$

where

$$f_s(w_{si}, d_i, q_i) = \sum_{i \in s} \frac{(w_{si} - d_i)^2}{d_i q_i}. \tag{4}$$

The equality in (3) called the calibration equation is the condition of model unbiasedness. The minimization of (4) allows to obtain design-consistent estimator. It is possible to prove under some conditions (see [1]) that the calibration estimator denoted by $\hat{\theta}^{CAL}$ obtained based on (3) but with different than (4) distance measure is asymptotically equivalent to the generalized regression estimator (1) in the sense that

$$N^{-1}\left(\hat{\theta}^{CAL} - \hat{\theta}^{GREG}\right) = O_p(n^{-1}). \tag{5}$$

Although (5) informs that the estimators are asymptotically equivalent, in [10] and [12] it is shown that the values of the estimators are very similar even for small sample sizes.

Weights (2) can be also obtained using instrumental-variable approach (see [2] p. 384). In this approach weights (2) are obtained based on the following equation:

$$w_{si} = d_i + (\sum_{i \in \Omega} \mathbf{x}_i - \sum_{i \in s} d_i \mathbf{x}_i)^T \left(\sum_{i \in s} c_i \mathbf{z}_i \mathbf{x}_i^T \right)^{-1} \mathbf{z}_i c_i, \tag{6}$$

where $\mathbf{z}_i$ is $p \times 1$ vector of instrumental variables, c_i are additional weights. To obtain (1) we use:

$$\mathbf{z}_i = \mathbf{x}_i \wedge c_i = d_i q_i. \tag{7}$$

The formula (6) means that weights w_{si} satisfy the calibration equation, but the choice of instrumental variables does not have to be based on the distance minimization approach as in (3).

Two design-consistent estimators of variance of (1) ([6] p. 15) are as follows:

$$\hat{D}_p^2(\hat{\theta}^{GREG}) = \sum_{j>i}^{n} \sum_{i}^{n} (\pi_i \pi_j - \pi_{ij}) \pi_{ij}^{-1} \left(d_i e_i - d_j e_j \right)^2, \tag{8}$$

$$\hat{D}_p^2(\hat{\theta}^{GREG}) = \sum_{j>i}^{n} \sum_{i}^{n} (\pi_i \pi_j - \pi_{ij}) \pi_{ij}^{-1} \left(d_i g_{si} e_i - d_j g_{sj} e_j \right)^2, \tag{9}$$

where $g_{si} = w_{si} d_i^{-1}$ and

$$e_i = y_i - \mathbf{x}_i^T \hat{\mathbf{B}}, \tag{10}$$

$$\hat{\mathbf{B}} = \left(\sum_{i \in s} d_i q_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i \in s} d_i q_i \mathbf{x}_i y_i. \tag{11}$$

3 Calibration estimators of subpopulation total

In the section we present two calibration estimators of subpopulation (domain) total. Let the population Ω be divided into D disjoint subpopulations (domains) denoted by Ω_d each of size N_d (where $d = 1, 2, ..., D$). Let the set of sampled elements from the *d*th domain be denoted by $s_d = s \cap \Omega_d$ and its size by n_d.

Let us introduce the first estimator of the *d*th domain total. The approximately design-unbiased (large *n* is required) and design-consistent (but if expected domain sample size is large as well) generalized regression estimator is given by ([6] pp. 17-18):

$$\hat{\theta}_d^{GREG} = \sum_{i \in s} w_{si} a_{id} y_i, \tag{12}$$

where $a_{id} = 1$ for $i \in \Omega_d$ and 0 otherwise. We can estimate the design-variance of (12) based on (9) where y_i is

replaced by $a_{id}y_i$. Hence, for $i \notin s_d$ we can obtain large negative residuals (10) given by $e_i = -\mathbf{x}_i^T \hat{\mathbf{B}}(a_{id}y_i)$ what may lead to inefficient design-variance estimator.

The second estimator of the *d*th domain total, called modified generalized regression estimator (MGREG), is approximately design-unbiased (large *n* is required). It is given by ([7]):

$$\hat{\theta}_d^{MGREG} = \sum_{i \in s_d} d_i y_i + \left(\sum_{i \in \Omega_d} \mathbf{x}_i - \sum_{i \in s_d} d_i \mathbf{x}_i \right)^T \hat{\mathbf{B}} \tag{13}$$

where $\hat{\mathbf{B}}$ is given by (11). To estimate design-variance of (13) we can use (8) where e_i are replaced by $a_{id}e_i$ assuming that the overall sample size is large even if the domain sample size is small ([6] p. 23).

Values of estimators (12) and (13) for all domains sum up to the value of estimator (1) of the population total: $\sum_{d=1}^{D} \hat{\theta}_d^{GREG} = \sum_{d=1}^{D} \hat{\theta}_d^{MGREG} = \hat{\theta}^{GREG}$. This property is called benchmarking. If in (3) we replace s by s_d, we obtain another GREG estimator of the domain total (see [6] p. 19) but without benchmarking property and which is not approximately design-unbiasedness (unless the domain size tends to infinity).

4 Correlation between variables and properties of estimators

Supporting estimation by using auxiliary variables is a common approach in survey sampling. Hence, it is necessary to review other authors' considerations in area of using auxiliary variables. We paid particular attention to the matter of dependence between variable under study and auxiliary variable as well as between auxiliary variables themselves.

In [1] there is considered a derivation of the ratio estimator. If in (2), $\mathbf{x}_i = x_i$ and we replace weights q_i by a $1/x_i$, then (1) will be the ratio estimator. The problem of correlation between variables in the case of the ratio estimator in simple random sampling without replacement is considered in ([5] pp. 165-166). According to [5] there ought to be a strong and positive correlation between the variable of interest and the auxiliary variable due to both the bias and the precision of the estimator. In [8] pp. 75-79 there are presented results of conducted simulation study, which objective was to verify what is the influence of using auxiliary variables and different correlation patterns on the bias and the variance of calibration estimators. In this study only positive dependence between the explanatory variable and the variable of interest was assumed. The problem of using auxiliary variables in the estimation of a finite population inter alia variance or covariance is undertaken in [11]. There are studied model-calibration, generalized difference estimators and also pseudoempirical likelihood methods. In the conducted simulation study, there were taken into consideration a few level of dependence between variables, but in all of them only positive relationship was assumed. In another simulation study (see [13]) investigating optimality of some calibration estimators also only positive dependence between the variable of interest (or its logarithm) and sum of auxiliary variables was presumed. In [3] there are proposed a few calibration estimators and then in a simulation study their properties are considered. Two models of generating data were considered, but in both of them there was a regression model with a positive slope, so only positive correlation between variables were studied.

This concise overview of literature shows that there is not enough research in the field of influence of dependence between a variable of interest and explanatory variables (especially in the case of negative relationship) and influence of correlation among auxiliary variables. That is why, in the next section we will present results of a simulation study where we investigate this topic.

5 Simulation studies

In subsections 5.1 and 5.2 we present results of design-based simulation studies (simple random sampling without replacement is used) conducted in R. In both simulation studies we analyze properties of the following estimators: GREGpop given by (1), GREGdom given by (12) and MGREGdom given by (13). We compute the relative biases of the estimators (in % of values of the real population/subpopulation totals), ratios of the mean square error (MSE) of calibration estimators to the MSE of Horvitz-Thompson estimator (HT) and relative biases of the estimators of the design-variances of calibration estimators (in % of design-variance of the estimator).

5.1 Real data

We use real data from Local Data Bank of The Central Statistical Office of Poland. The considered dataset consists of $N = 379$ Polish counties (the capital city - Warsaw was omitted as an outlier). We use 3 variables: the variable

of interest (y) which is the number of employed persons and two auxiliary variables: the number of renovation dwellings (x_1), and the amount of permits granted for construction (x_2). The population is divided into six disjoint domains - regions in Poland (NUTS 1). We select auxiliary variables in the way to achieve high level of dependency with the variable of interest (the correlation coefficients are 0.85 and 0.92, respectively). Unfortunately, there is also a problem of high correlation between auxiliary variables, the correlation coefficient in that case is almost 0.7.

Firstly, on Figure 1 we analyze the influence of using both or one particular auxiliary variable on relative biases of calibration estimators. In all 3 cases the obtained results are quite similar, but the smallest biases we obtain in the case where only x_1 is used. In the right graph in Figure 1 we present ratios of MSE of calibration estimators for domains (population) and HT for domains (population). The aim of this comparison is to show a decrease in MSE using calibration estimators instead of HT estimator. For MGREGdom and GREGpop all obtained ratios take values lower than 1, therefore we can conclude that for these estimators their MSE is smaller comparing to the MSE of HT estimator. For GREGdom estimator the ratios are usually above 1, so it means that in most of the cases there is no decrease in MSE using GREGdom instead of HT.

Secondly, we study properties of estimators of design-variance of calibration estimators. Figure 2 displays its relative biases. With regard to the results obtained in domains, it can be observed that relative biases are from -40% to 30% and also that for MGREGdom we usually received lower values of the bias. When it comes to the relative biases of design variance of GREGpop, in the case of using two auxiliary variables we obtain very high value of bias, it is almost 300%. In the case of using only one of the auxiliary variables, we receive results similar to the results for relative biases in domains.

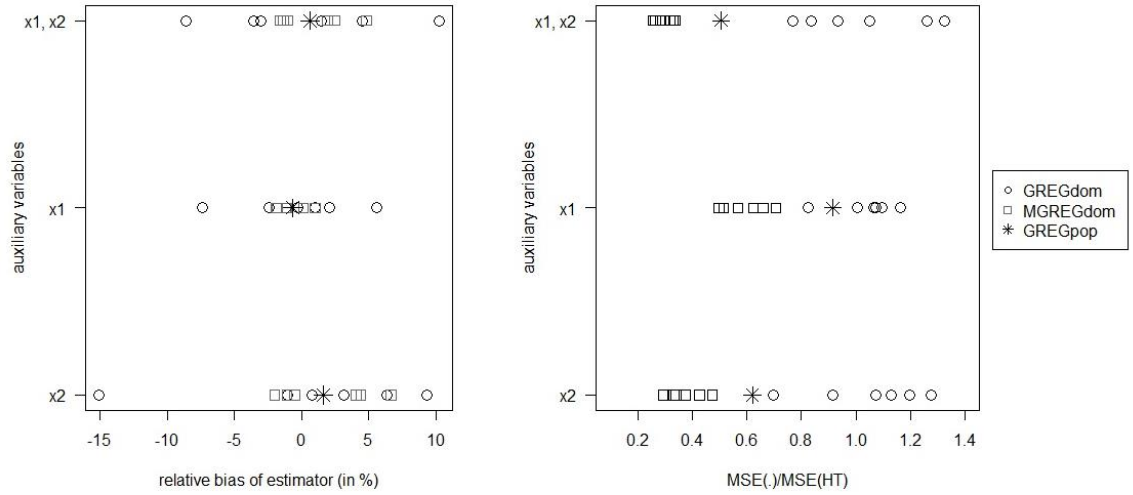


Figure 1 Relatives biases of calibration estimators (on the left) and ratios of the MSE of calibration estimators and the MSE of HT estimator (on the right) for the real data

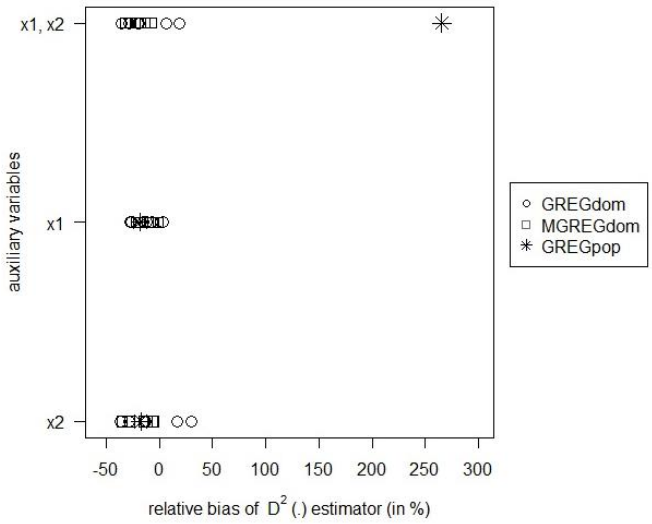


Figure 2 Relatives biases of design-variance of calibration estimators for the real data

5.2 Artificial data

We generate values the variable under study and two auxiliary variables based on Gaussian copulas with gamma marginal distributions with the mean and the variance of the respective real variables presented in the previous subsection and with different correlation matrices. Moreover, for all generated data we assumed the same value of the third standardized moment which equals 1. The reason is fact that generating data with the higher level of asymmetry gives in our case parameters of generated data significantly different from the assumed values.

We consider eleven cases of the correlation between variables of interest and auxiliary variables (see OY axis on Figure 3). In 9 assumed correlation matrices we use two auxiliary variables. We study the following cases: both explanatory variables are highly positively correlated with the variable of interest, both explanatory variables are highly negatively correlated with the variable of interest and in the last case we assume that one of the auxiliary variable is highly positively correlated and the second one is negatively correlated with the variable under study. These three models are crossed with three models of different levels of dependency between auxiliary variables. We assumed weak positive correlation between auxiliary variables, weak negative correlation between auxiliary variables and moderate positive or negative correlation between the variables. Moreover, we study two cases where only one auxiliary variable is used: positively or negatively correlated with the variable under study. All of the assumed cases of values of correlation coefficients (see Figure 3) give positive semi-definite correlation matrices.

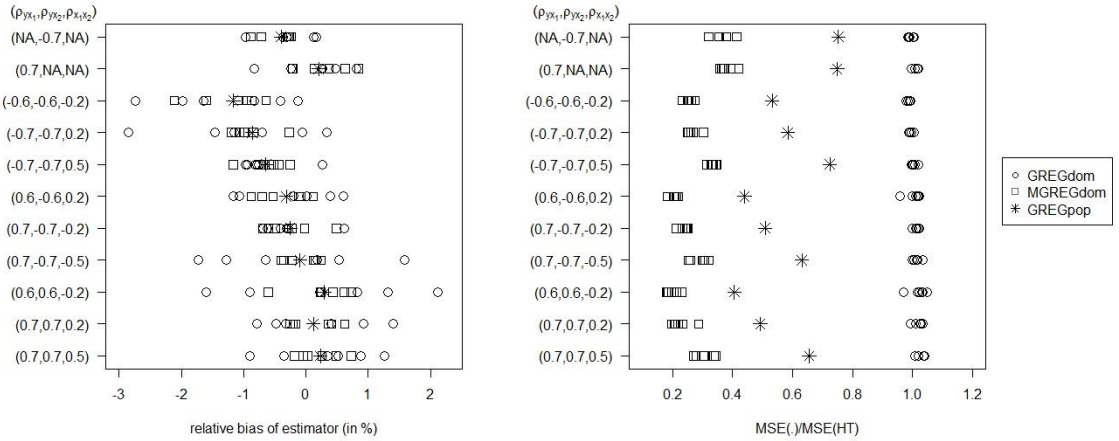


Figure 3 Relatives biases of calibration estimators (on the left) and ratios of the MSE of calibration estimators and the MSE of HT estimator (on the right) for the artificial data

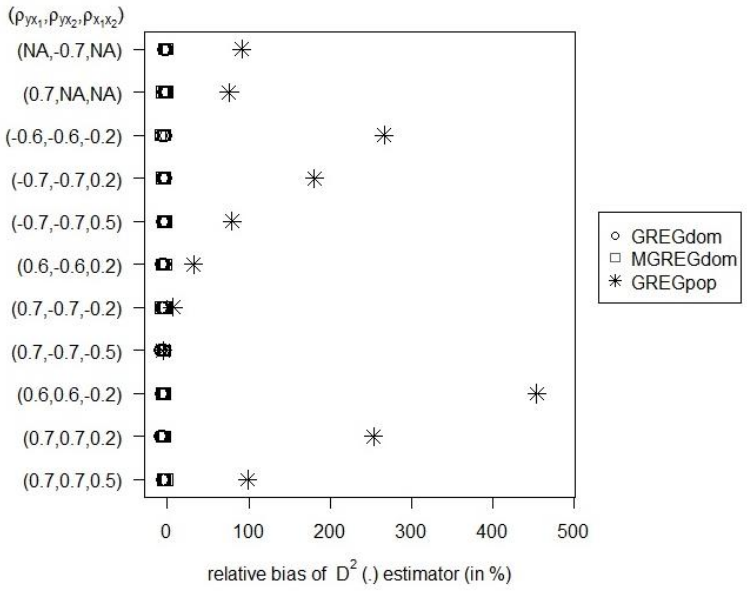


Figure 4 Relatives biases of design-variance of calibration estimators for the artificial data

In the first step we analyze relative biases of the calibration estimators. They are shown on the left side of Figure 3. For all considered cases of the correlation matrices we received similar results. In each case, absolute values of relative biases are below 3%. It is worth noting that for the artificial data we obtained significantly smaller values of the relative biases than for the real data. On the right side of Figure 3 we show the comparison of the MSE of the calibration estimators to the MSE of HT estimator. For GREGdom estimator we obtained ratios, which are in each case very close to 1. As regards the MSE of MGREGdom estimator the lowest values of ratios (so the largest gain comparing to the MSE of HT estimator) we received for cases: 2, 3 and 6 (counting from the bottom on OY axis). In those cases the level of dependence between auxiliary variables was low – slightly positive or negative. The largest values of MSE ratios (but still smaller than 1) were obtained for correlation matrices with only one auxiliary variable and for matrix with auxiliary variables correlated negatively with the variable under study where auxiliary variables were moderately positively correlated with each other. To sum up, in each case the MSE of MGREGdom was lower than the MSE of HT in domains. Analogous results were obtained comparing the MSE of GREGpop to the MSE of HT estimator for the whole population.

Secondly, we study properties of estimators of design-variance of calibration estimators. Received results are shown in Figure 4. For domain estimators GREGdom and MGREGdom obtained biases were low - from -10% to 10%. In case of GREGpop we have greater diversity. For cases with two auxiliary variables: one positively and one negatively correlated with the variable under study obtained relative biases were low – their absolute values were below 10%. But for some cases, we obtained huge biases, for instance for case 3 (counting from the bottom on OY axis) with both auxiliary variables positively correlated with the variable under study, obtained relative bias is over 400%. We suppose that such huge biases can be result of slightly weaker dependence with the variable under study.

6 Conclusion

We analyze in the simulation studies how many auxiliary variables should be used and which level of dependency between auxiliary variables and between auxiliary variables and the variable under study allows to increase accuracy of calibration estimators and give low values of estimators of their design-variance. Conclusions for these two issues are not the same. It turns out that in order to decrease the MSE of the calibration estimators we should use rather two auxiliary variables with weak level of dependence between them. As regards properties of estimators of design-variance in order to get the lowest biases it is recommended to use one auxiliary variable positively correlated and the second negatively with the variable under study or to use only one auxiliary variable, because in some cases with two variables huge biases could be obtained.

References

[1] Deville J. C., and Särndal C. E.: Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87** (1992), 376-382.

[2] Estevao, V.M., and Särndal, C.E.: A Functional Form Approach to Calibration. *Journal of Official Statistics* **16** (2000), 379-399.

[3] Kim, J.K., and Park M.: Calibration Estimation in Survey Sampling. *International Statistical Review* **78** (2010), 21-39.

[4] Krzciuk, M.K.: On the design accuracy of Royall's predictor of domain total for longitudinal data. In: *32nd International Conference Mathematical Methods in Economics MME 2014. Conference Proceedings.* (Talašová, J., Stoklasa, J. and Talášek, T., eds.), Palacky Univerisity, Olomouc, 2014, 512-517.

[5] Pawłowski, Z.: *Wstęp do statystycznej metody reprezentacyjnej. Państwowe Wydawnictwo Naukowe*, Warszawa, 1972.

[6] Rao, J. N. K, and Molina I.: *Small area estimation*. Second edition. John Wiley and Sons, Hoboken, New Jersey, 2015.

[7] Särndal, C. E.: Frameworks for Inference in Survey Sampling with Applications to Small Area Estimation and Adjustment for Nonresponse. *Bulletin of the International Statistical Institute* **49** (1981), 494-513.

[8] Särndal, C. E., and Lundström, S.: *Estimation in Surveys with Nonresponse*. John Wiley and Sons, Ltd, 2005.

[9] Särndal, C.E., Swensson, B., and Wretman J.: *Model Assisted Survey Sampling*. Springer-Verlag, New York, 1992.

[10] Singh, A.C., and Mohl, C.A.: Understanding calibration estimators in survey sampling. *Survey Methodology* **22** (1996), 107-115.

[11] Sitter, R.R., and Wu, C.: Efficient Estimation of Quadratic Finite Population Functions in the Presence of Auxiliary Information. *Journal of the American Statistical Association* **97** (2002), 535-543.

[12] Stukel, D.M., Hidiroglou, M.A., and Särndal, C.E.: Variance estimation for calibration estimators: A comparison of jackknifing versus Taylor linearization. *Survey Methodology* **22** (1996), 177-125.

[13] Wu, C.: Optimal calibration estimators in survey sampling. *Biometrika* **90** (2003), 937-951.

Identification of bankruptcy factors for engineering companies in the EU

Michaela Staňková[1], David Hampel[2]

Abstract. This article focuses on the possibility of identification of factors leading to bankruptcy of the companies from engineering sector in EU. We deal with 953 small and medium-sized engineering companies in the EU from which 51 companies bankrupted in 2014. The data are collected from the Amadeus database of Bureau van Dijk. At first, we use factor analysis for reduction of the data dimension. Several meaningful results with different number of factors are found. We use the method of cluster analysis (with different settings) in order to identify the group of active and bankrupted companies; this analysis is based on original financial data and on artificial factors. Both steps are repeated for the data coming from one, two and three years before bankruptcy. Influence of data dimension reduction on cluster analysis results is discussed and final recommendations for further analyses are given.

Keywords: factor analysis, cluster analysis, bankruptcy prediction.

JEL Classification: C38
AMS Classification: 62M25, 62H30

1 Introduction

Evaluating the financial situation of company is a very important topic for decades in the academic or practical field of corporate finance. Knowledge of the company's financial situation is important not only for the shareholders, creditors and top management of the company in their decision-making, but also for the many other subjects. Because the financial distress of the company (especially for a larger number of companies over the same period) brings serious problems such as unemployment. Therefore, there is a constant demand for ever more accurate and stable tool for predicting the company's financial situation.

Since the second half of the last century, when the evaluation of the financial status of the company began to focus more attention, a myriad of models were developed to predict the financial situation of a company. Although the models differ by the method used, all the models evaluating the financial situation of the company are based on the premise that companies have symptoms typical of bankruptcy for some time before this status. In all these models, financial ratios play a significant role.

Based on the company's financial statements, it is possible to define many financial indicators that could indicate the bankruptcy of the company. From a practical point of view, the models consist of only a few indicators. Among the world's most well-known models belongs the Altman Z-Score, see [2]. This model is based on five financial ratios. Beaver in [5] also chose five indicators. When using the Taffer model introduced in [1], only four ratios are used. Beerman's function describe in [6] includes a total of 10 financial ratios. Differences can be found not only in the number of used financial ratios but also in their own construction. Beerman, like Beaver, used for example the cash flow/total liabilities. But the Altman or Taffler model does not contain any financial ratios that use the amount of cash flow. It is possible to say, that there is no general agreement on the optimal selection of ratios.

This article focuses on the process of identifying significant financial indicators that could be used to compile a model for bankruptcy prediction of company. It is equally important to choose the appropriate method for selecting the most important variables. For our analysis, factor analysis (FA) and cluster analysis (CA) are selected. An advantage of the factor analysis is definitely its feature to reduce the number of observed variables connecting them into latent variables that explain the correlated content of observed variables. On the other hand, cluster analysis explores the types of data relations instead of their correlations as stated in [11]. Cluster analysis is focused on the partitioning of similar objects into meaningful classes when both the number of classes and the composition of the classes are to be determined, see [3]. Recent studies used CA and FA in order to identification significant factor from the field of bankruptcy assessment are [4], [7], [8], [9] and [10].

[1]Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, xstanko7@node.mendelu.cz
[2] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, david.hampel.uso@mendelu.cz

2 Material and Methods

The financial (annual accounting) data are collected from the Amadeus database of Bureau van Dijk. We deal with 953 engineering companies in the EU from which 51 companies bankrupted in 2014. The models are constructed separately for the period of one, two and three years before the company's bankruptcy. To achieve more homogeneous data set, we include only small and medium-sized engineering companies. With regard to the financial theory, we choose following 19 indicators representing the four most common groups of financial indicators (i.e. solvency ratios, profitability ratios, liquidity ratios and turnover ratios). In Table 1, there are financial variables used in our analysis.

Working capital per employee (EUR)	Net working capital (EUR)	Collection period (days)	Credit period (days)
Solvency ratio (%)	Debt to equity ratio	Debt ratio (%)	Interest cover
Net assets turnover	Stock turnover	Debt repayment period	Fixed asset to equity capital ratio
Current ratio	Quick ratio	Cash ratio	Cash flow liquidity
EBIT margin (%)	Return on assets (%)	Profit margin (%)	

Table 1 Overview of financial ratios used in the analysis

A principal component analysis (PCA) together with FA was performed on a set of selected variables. PCA allowed us to make the necessary evaluations using the artificial variables that explain the same variability as the original values of financial ratios. Thanks to the information from the PCA, a reduction of the original data dimension through FA is made without significant loss of information. Scree plot was used to find a suitable number of factors; we will also check that the selected factors together explain at least 80% of the original variability. FA was made with three different settings – without rotation along with varimax and quartimax axis rotation.

The CA was used to describe differences and relationships within selected group of companies. Agglomerative clustering has been chosen for our purposes. Combined with the financial ratios of the company, it is possible to use the CA for classification of companies according to predefined parameters. Consequently it can be used for evaluating the performance, creditworthiness or financial distress (also bankruptcy) of companies. Even with CA, we use different settings to generalize and compare results. As a pairwise distance between pairs of objects we choose Euclidean distance and standardized Euclidean distance. For computing distance between clusters we used: Shortest distance, Furthest distance, Average distance and Ward distance.

CA will be made via all 19 financial ratios and also based on identified artificial factors and repeated for the data coming from one, two and three years before bankruptcy. The identified clusters will be categorized as “active” or “bankruptcy” by the proportion of companies in them. We are guided by the simple rule that if the proportion of bankruptcy companies in a cluster is at least 50%, cluster is marked as bankruptcy. All calculations were performed in computational system Matlab.

3 Results

Using the scree plot for each period, we detected that it would be appropriate to work with two to a maximum of six factors. Moreover, the first two components together explained over 80% of the original variability in most cases. FA without rotation and with varimax or quartimax axis rotation gave us the same composition of the factors. For both axis rotations, the correlation between the detected factors was only slightly increased. After performing the FA, we found that it would be appropriate to work at most with the first five factors. The sixth factor was composed of financial ratios that have only a weak correlation to them. In following tables, there are identified factors for each monitored period.

In Table 2, we see that several groups of ratios have been identified using FA for the period of three years before bankruptcy. One factor is composed of indicators that measure the ability of a company to meet its short-term obligations. For this reason it was named as Liquidity. Another factor has been labeled Profitability because it includes indicators that work with some form of profit. We have also identified a factor indicating the Level of indebtedness as it reflects the capital structure of companies. The factor linking the Collection period and the Credit period was marked as Customer-supplier relationships, because company has a certain level of liquidity with the right set-up of these relationships.

Number of factors	Name of the factor	Financial ratios forming the factor
2	Liquidity	Current ratio, Quick ratio, Cash ratio, Cash flow liquidity
	Profitability	Return on assets, EBIT margin, Profit margin
3	Profitability	Return on assets, EBIT margin, Profit margin
	Liquidity	Current ratio, Quick ratio, Cash ratio, Cash flow liquidity
	Level of indebtedness	Solvency ratio, Debt ratio
4	Liquidity	Current ratio, Quick ratio, Cash ratio, Cash flow liquidity
	Profitability	Return on assets, EBIT margin, Profit margin
	Level of indebtedness	Solvency ratio, Debt ratio
	Customer-supplier relationships	Collection period, Credit period
5	Profitability	Return on assets, EBIT margin, Profit margin
	Level of indebtedness	Solvency ratio, Debt ratio
	Amount of current liabilities	Quick ratio, Cash flow liquidity, Working capital per employee
	Current ratio	Current ratio
	Customer-supplier relationships	Collection period, Credit period

Table 2 Individual factors for the period of three years before bankruptcy

Group of five factors have already begun to show a declining level of linkages between financial ratios. This fact was reflected in a less meaningful grouping of factors. Current ratio eventually remained as one whole factor. The original Liquidity Factor has been split up. The value of the Working capital per employee has appeared, that is why we made up a new name of this factor Amount of current liabilities.

Table 3 lists the identified factors for period of two years before bankruptcy. In this period (apart from the aforementioned factors from the previous period) a relatively disparate factor was arose, called Amount of sales assets. We assume that bankrupt companies were selling their assets at this time. This fact was reflected in assets and cash flow as well. In the case of identifying a group of five factors, Amount of sales assets factor was split and two new factors emerged. One consisting only of Solvency ratio and second one denoted as Sale of assets.

Number of factors	Name of the factor	Financial ratios forming the factor
2	Liquidity	Current ratio, Quick ratio, Cash ratio
	Profitability	EBIT margin, Profit margin
3	Liquidity	Current ratio, Quick ratio, Cash ratio
	Profitability	EBIT margin, Profit margin
	Customer-supplier relationships	Collection period, Credit period
4	Amount of sales assets	Solvency ratio, Debt ratio, Return on assets, Cash flow liquidity
	Liquidity	Current ratio, Quick ratio, Cash ratio
	Profitability	EBIT margin, Profit margin
	Customer-supplier relationships	Collection period, Credit period
5	Liquidity	Current ratio, Quick ratio, Cash ratio
	Sale of assets	Return on assets, Debt ratio, Cash flow liquidity
	Profitability	EBIT margin, Profit margin
	Customer-supplier relationships	Collection period, Credit period
	Solvency ratio	Solvency ratio

Table 3 Individual factors for the period of two years before bankruptcy

Even in the period of one year before bankruptcy, see Table 4, new combinations of financial indicators were formed. In the first place (in all identified groups) there is a factor pointing to Ability to trade. The latest newly-defined factor reflects the amount of inventory. Bankruptcy companies also significantly reduced the amount of inventory in this period. That is why we called this factor Sale of goods in stock.

Number of factors	Name of the factor	Financial ratios forming the factor
2	Ability to trade	EBIT margin, Profit margin, Collection period, Credit period
	Liquidity	Current ratio, Quick ratio, Cash ratio
3	Ability to trade	EBIT margin, Profit margin, Collection period, Credit period
	Liquidity	Current ratio, Quick ratio, Cash ratio
	Sale of assets	Return on assets, Debt ratio, Cash flow liquidity
4	Ability to trade	EBIT margin, Profit margin, Collection period, Credit period
	Liquidity	Current ratio, Quick ratio, Cash ratio
	Sale of assets	Return on assets, Debt ratio, Cash flow liquidity
	Level of indebtedness	Solvency ratio, Debt ratio
5	Ability to trade	EBIT margin, Profit margin, Collection period, Credit period
	Liquidity	Current ratio, Quick ratio, Cash ratio
	Sale of assets	Return on assets, Debt ratio, Cash flow liquidity
	Level of indebtedness	Solvency ratio, Debt ratio
	Sale of goods in stock	Stock turnover, Working capital per employee

Table 4 Individual factors for the period of one year before bankruptcy

In the second step, the CA was performed. For each of period we made CA via all the obtained values (i.e. 19 financial indicators) and via the chosen factors. Regardless of the chosen distance between pairs of objects, the Ward method appears to be the most appropriate method for the distance between clusters in all three observed periods. This method gave the most transparent graphical outputs. Table 5 shows results of CA with Ward distance setting using the data from period three years before bankruptcy. Based on these results, we can conclude that the differentiation of bankruptcy companies was difficult for this period. The best clustering results were achieved through three and four factors, where four to five clusters were identified. Only one of these clusters was then flagged as bankrupt.

Input data	Distance between objects	Number of clusters	Number of clusters identified as bankruptcy	Bankruptcy companies correctly classified (%)
All 19 ratios	Euclidean	8	0	0
2 factors	Euclidean	4	0	0
3 factors	Euclidean	5	1	11.76
4 factors	Euclidean	4	1	9.80
5 factors	Euclidean	6	0	0
All 19 ratios	std. Euclidean	8	0	0
2 factors	std. Euclidean	4	0	0
3 factors	std. Euclidean	4	0	0
4 factors	std. Euclidean	4	1	13.73
5 factors	std. Euclidean	4	0	0

Table 5 Individual factors for the period of three years before bankruptcy

In the period of two years before bankruptcy, the best results were achieved with slightly different settings. The most appropriate choice here is CA based on four or five factors, using the Ward method with Euclidean distance between objects. In Table 6, there are results of CA via Ward method using the data from period three years before bankruptcy.

Input data	Distance between objects	Number of using clusters	Number of clusters identified as bankruptcy	Bankruptcy companies correctly classified (%)
All 19 ratios	Euclidean	7	1	1.96
2 factors	Euclidean	5	1	27.45
3 factors	Euclidean	6	1	21.57
4 factors	Euclidean	5	1	27.45
5 factors	Euclidean	6	1	27.45
All 19 ratios	std. Euclidean	7	1	11.76
2 factors	std. Euclidean	6	1	11.76
3 factors	std. Euclidean	7	3	21.57
4 factors	std. Euclidean	7	3	19.61
5 factors	std. Euclidean	7	3	19.61

Table 6 Individual factors for the period of two years before bankruptcy

In the last reference period, the ideal number of factors is four. Using this number of factors together with the Ward method and the standardized Euclidean distance between objects, the CA results were most accurate. In Table 7, there are results of CA with Ward distance using the data from period one year before bankruptcy.

Input data	Distance between objects	Number of using clusters	Number of clusters identified as bankruptcy	Bankruptcy companies correctly classified (%)
All 19 ratios	Euclidean	4	0	0,00
2 factors	Euclidean	6	4	29.41
3 factors	Euclidean	5	4	41.18
4 factors	Euclidean	5	4	41.18
5 factors	Euclidean	6	3	39.22
All 19 ratios	std. Euclidean	4	1	15.69
2 factors	std. Euclidean	5	2	45.10
3 factors	std. Euclidean	5	2	62.75
4 factors	std. Euclidean	5	2	64.71
5 factors	std. Euclidean	6	2	60.78

Table 7 Individual factors for the period of one year before bankruptcy

4 Discussion

In general, CA (with different settings) in all three monitored seasons has shown the suitability of using only a few factors instead of using all financial ratios. In the case of clustering on the basis of all the obtained financial data, almost all of the bankruptcy companies got into one large cluster. However, this cluster was shared with hundreds of other active companies. Using a standardized Euclidean distance between objects, the number of these active companies in a given cluster changed significantly. It was a change of roughly half the cluster size. But even with this setting, the classification successes of CA based on all financial indicators did not reach the level of the version based on the identified factors.

In the period of one year before bankruptcy, we were able to correctly identify almost 65% of bankruptcy companies with CA. Although with the increasing time period since the bankruptcy year the prediction ability of this technique is declining. In the period of three years before bankruptcy, we were able to correctly identify less than 14% of bankruptcy companies through this procedure.

Based on the empirical results (with different CA settings), the most accuracy classification appears to be via four factors. These four factors are composed of 11 financial ratios representing all four of the most frequently mentioned groups of ratios. However, the importance of these financial ratios has changed considerably over time. For example, at first the most variability was covered by the Liquidity factor. Since the second period, the financial ratios constituting this factor have reached the second position. As the bankruptcy is approaching, the significance of the turnover ratios (i.e., Collection period and Credit period) is enhanced.

The identified factors used in our article for CA purposes could also be used for other classification algorithms such as support vectors machines or decision trees. It is therefore possible to focus on financial distress prediction models using these identified factors.

5 Conclusion

In our paper, we tried to identification of factors leading to bankruptcy of the companies from engineering sector in EU. The data set includes 902 active companies and 51 companies which have gone bankrupt in 2014. At first, we use FA for reduction of the data dimension. Regardless of the settings used, it was found that it is appropriate to use at most the first five factors. Factors were identified for each of the monitored periods. These factors include financial indicators representing all four bases groups of financial ratios (i.e. solvency ratios, profitability ratios, liquidity ratios and turnover ratios).

We use the method of CA (with different settings) in order to identify the group of active and bankrupted companies. Based on empirical results, we can say that better results are achieved if financial data before CA is processed by FA. In general, we can recommend the Ward method for the distance between clusters. In all three monitored periods, the choice of CA based on four factors seems appropriate. The choice of standardization or non-standardization of the Euclidean distance between objects is then ambiguous. In the period of two years before bankruptcy, in the case of Euclidean distance, much better results were achieved, but for the other two periods, it was preferable not to use standardization.

Acknowledgements

This article was supported by the grant No. PEF/DP/2017027 of the Grant Agency IGA PEF MENDELU.

References

[1] Agarwal, V., Taffler, R.. J.: Twenty-five years of the Taffler Z score model: does it really have predictive ability? *Accounting & Business Research*, **37** (2007), 285-300.
[2] Altman, E.: Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* **23** (1968), 589-609.
[3] Anderberg, M. R. *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks*. Academic press, New York, 2014.
[4] Armeanu, S. D., Vintila, G., Moscalu, M., Filipescu, M. O., Lazar, P.: Using Quantitative Data Analysis Techniques for Bankruptcy Risk Estimation for Corporations. *Theoretical and Applied Economics* **19** (2012), 97–112.
[5] Beaver, W.: Financial Ratios as Predictors of Failure. *The Journal of Accounting Research* **4** (1966), 71-102.
[6] Beerman, K.: Possible Ways to Predict Capital Losses with Annual Financial Statements. Düsseldorf, Germany, 1976.
[7] Chen, M. Y.: Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications* **38** (2011), 11261–11272.
[8] Klepáč, V. Hampel, D.: Predicting financial distress of agriculture companies in EU. *Agricultural Economics* (2017), Online First: DOI:10.17221/374/2015-AGRICECON.
[9] Klepáč, V. Hampel, D.: Prediction of Bankruptcy with SVM Classifiers Among Retail Business Companies in EU. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **64** (2016), 627–634.
[10] Marinč, V., Rant, V.: A cross-country analysis of bank bankruptcy regimes. *Journal of Financial Stability* **13** (2014), 145–150.
[11] Sobíšek, L., Řezanková, H.: Srovnání metod pro redukci dimensionality aplikovaných na ordinální proměnné. *Acta Oeconomica Pragensia* **1** (2011), 3–19.

Wave relations of exchange rates in binary-temporal representation

Michał Dominik Stasiak[1]

Abstract. An exchange rate trajectory can be given in a binary-temporal representation. Basis of this representation is the course discretization, in which to every change of exchange rate equal to a given discretization unit two parameters are assigned, that is binary value corresponding to the direction of trajectory change and its duration. Statistical analysis performed by the Author confirms relationships between previous changes and their duration, and the future direction of changes. One of the fundamental methods of technical analysis is wave analysis. In the presented article some algorithms were given, allowing for detecting a start, end and duration of a given wave in binary-temporal representation of exchange rate trajectory.
A detailed statistical analysis of wave parameters for AUD/NZD pair in five-year period (2012-2017) was also performed. The main part of this article consist of research dedicated to ascertain the influence of type and actual wave parameters on the probability of future course change. Obtained results indicate a high predicative value of proposed model. The binarization process of exchange rate course and appointment of wave parameters were performed based on specialistic software, developed in MQL4 and C++.
Keywords: Foreign exchange market, high frequency econometric, technical analysis, currency market investment decision support, modelling of currency exchange rates.

JEL Classification: F31, G11, G14, C49
AMS Classification: 91G70, 62P20

1 Introduction

Most technical analysis methods use candlestick charts [6,9,11]. Candlestick representation is also common in broker platforms such as MetaTrader or JForex. Sadly, this representation causes a loss of important information regarding the exchange rate variability "inside" the candle. This observation matters greatly when considering small transactions with range of, for example, 15 pips. In this case, even one minute candlesticks do not prevent losing significant information about the course trajectory, since in times of high activity even in one-minute period a lot of such changes can occur. Disadvantages of candlestick representation are especially visible in creating prediction models in High Frequency Trading (HFT), in which hundreds and thousands of transactions are made. The transactions have a small, few-pips range and the profit of an investor is ensured by statistical prediction of transaction direction [1]. In this context, it is highly more advisable to use binary [12] or binary-temporal representation [13], in which to each change of given range a binary value is being assigned.

Wave analysis is one of the basic technical analysis methods. Relations between parameters of ensuing waves were presented, among others, in Elliott's theory [5]. However, those dependences are of a highly general character and thus, regarding the necessity of visual detection of a wave, building HFT models based on wave analysis is very difficult. In the following article, a wave detection algorithm for an exchange rate in a binary-temporal representation is presented. A dependency analysis for probabilities of future change direction of given range and wave parameters is also performed for a currency pair.

2 Binary-temporal representation

Most of the technical analysis methods use candlestick representation and majority of indicators are appointed based on one parameter of a candle [6,9,11]. Course trajectory analysis in candlestick representation leads to a loss of significant information about order and number of small changes "inside" the candle. In case of small-range transactions – occurring inside a single candle – it is often difficult to specify which of the parameters: Take Profit (TP) or Stop Loss (SL) was reached first and, as a consequence, if the investor had achieved a profit

[1] Poznań University of Economics and Business, Department of Investment and Real Estate, Niepodległości 10, 61-875 Poznań, Poland e-mail: michal.stasiak@ue.poznan.pl

or a loss. Analysis of candlestick charts, as well as many other methods of technical analysis, allow only for determination of the future changes direction, without specifying their range [6,9,11]. This rises the problem of proper appointment of TP and SL parameters. In consequence, course trajectory modelling in candlestick representation is inaccurate and unreliable. Tick data analysis is also not very effective because of the noise phenomenon, that is a lot of small changes of a random character [7,8]. In the article, a binary-temporal representation is proposed for modelling the course trajectory. Methods allowing a precise analysis of changes with given range usually use discretization. The exchange rate discretization was first used at the beginning of XX century, in creating and analyzing charts in so called point-symbolic method [4]. The method, despite being more accurate than others, was soon replaced by candlestick chart analysis methods and was no longer developed. A disadvantage of both point-symbolic method and binary representation [12] is the lack of information about the duration of a change. The main idea of this representation lies in appointing a binary value to each course change equal to a discretization unit, i.e. 1 for an increase and 0 for a decrease, as well as the duration of a change given in minutes [13]. Figure 1 shows an example of tick data conversion into binary-temporal representation.

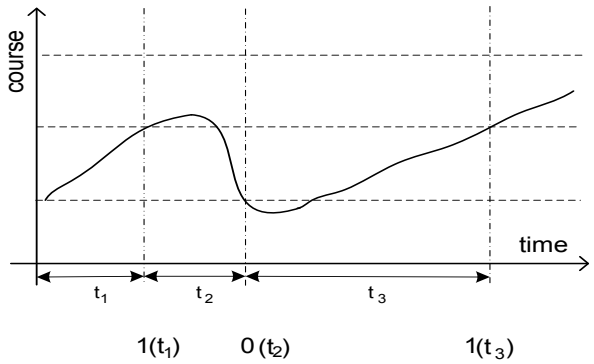


Figure 1. Example of tick data conversion with a binary-temporal representation

Binary-temporal representation can be used for modelling the course trajectory in HFT systems. Let us assume, that HFT system makes, for example, purchase transactions with parameters TP and SL both placed one discretization unit above and below the price. Then, the profit probability is equal to the probability of value 1 occurring in binary-temporal representation. Similarly, loss probability is equal to the probability of value 0. Therefore, choosing effective methods for predicting ensuing binary values leads to a creation of HFT systems with positive return rate. The quality of models in binary representations depends on the assumed discretization unit. The lower the value, the more accurate are the results of course modelling. On the other hand, too small discretization unit can lead to recording redundant random fluctuations [7,8]. Additionally, taking into account such practical aspects as, for example, spread level (ca. 2 pips for analyzed pair), etc., using a 15-pips discretization unit seems to be optimal. Detailed statistical analysis of curse trajectory presented in a binary-temporal representation showed visible dependences between the historical changes and the direction of future changes of the analyzed currency pair [13].

3 Wave analysis

Wave analysis is one of the most popular methods of technical analysis. It is based on an assumption that mutually dependent waves occur on the market. Wave structure of a market was escribed in detail in Elliott's theory [5]. The existence of waves is commonly accepted and can be explained by e.g. investors' reaction to new, important information [10]. For instance, after announcing important macroeconomic data investors make decisions based on this information for a period of time after the actual announcement. As a result, a wave occurs. Still, determining relations between parameters of ensuing waves causes some doubts. Most of the wave analysis methods, e.g. formation analysis or Elliott's theory, assume visual detection of waves on a candlestick chart [2,3,8]. Besides the information loss in candlestick charts, the effectiveness of the method relies greatly on analyst's subjective assessment. Without precisely described rules of wave detection, reliable statistical analysis and objective verification of potential wave dependences is not possible. In the Figure 2 we can see two (of many) possibilities of wave detection pointed by an investor or analyst for given exchange rate. In this article a new wave detection algorithm is proposed for a currency pair in binary-temporal representation, aiming to avoid above stated problem. It allows for an increase and decrease wave detection and objectively appointing their parameters.

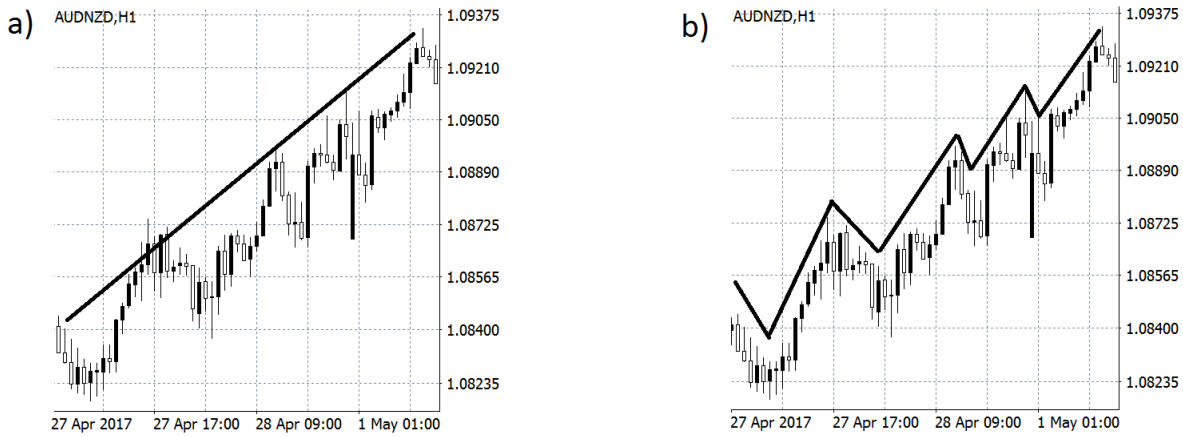


Figure 2. Different ways of visual wave detection

3.1 Wave detection algorithm in binary-temporal representation

Wave Detection Algorithm in Binary-Temporal Representation (WDABTR) allows for detecting the wave type and appointing its parameters in binary-temporal representation. It determines if the current wave W_i is increasing (denoted 1) or decreasing (0) for i-th course trajectory change of one discretization unit, appoints its duration (T_i) and length (L_i.) The wave length is described by a number of changes counted from its first registration, and duration, being a sum of durations of single changes creating a wave, is given in minutes. Algorithm detects the type of a wave based on an analysis of average number of increases and decreases in a given historical period. It is also assumed that the current wave is continued as long as a new one does start.

Wave detection algorithm uses the following parameters:

l – number of analyzed historical changes,
p – wave detection threshold.

In the algorithm:

u – average number of increases in l changes starting an increase wave;

$$u = \frac{\sum_{i=1}^{l} z_i}{l}, \tag{1}$$

d – average number of decreases in l changes starting a decrease wave;

$$d = \frac{\sum_{i=1}^{l} |z_i - 1|}{l}, \tag{2}$$

where z_i is an i-th change in a binary representation. We have a following relation between w and s:

$$u + d = 1 \tag{3}$$

T_f – duration of current wave after i changes.

$$T_f = \sum_{k=i}^{i} t_k, \tag{4}$$

where t_i is duration of i-th change,

D_f –length of current wave after i changes.

$$D_f = i, \tag{5}$$

Accuracy and frequency of detected waves are dependent on assumed parameters l and p. Their value also dictates the character of dependences between ensuing waves. A simplified block diagram of WDABTR performance is presented in the Figure 3.

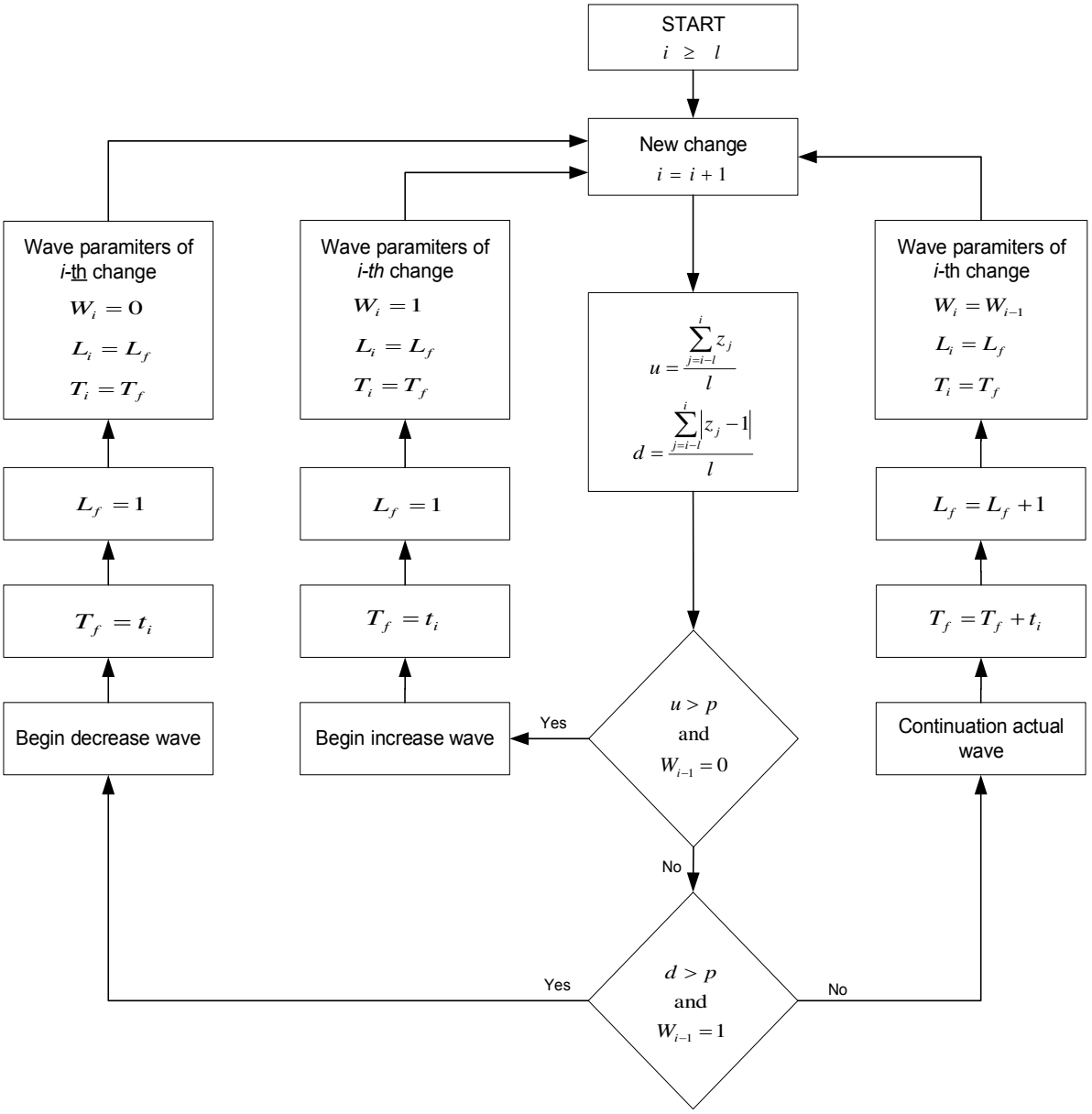


Figure 3. WDABTR block diagram

3.2 Wave representation

Based on the binary-temporal representation and the WDABTR, exchange rate can be given in a wave representation. Then, to each course change of one discretization unit length, three parameters are assigned. First parameter is connected with the wave type W_i. Second one describes the current wave length L_i given as a number of changes which occurred since the wave registration. The third one is the current wave duration T_i, given in minutes. Figure 4 presents an example of course trajectory notation in wave representation. It allows for a clear description of current market situation. The notation simplifies course modelling and prediction model construction, which main goal is to reliably assess the probabilities of future changes direction based on wave dependences in previously registered changes.

i	·	75	76	77	78	79	·
z_i	·	1	1	1	0	1	·
W_i	·	0	0	1	1	1	·
L_i	·	3	4	1	2	3	·
T_i	·	8	11	2	9	13	·

Figure 4. Wave representation of an exchange rate

4 Wave dependences analysis

Let us now consider wave representation obtained based on binary-temporal representation with 15-pips discretization unit, and WDABTR with parameters l=4 and p=0.61. We take a 5-year historical data for ADU/NZD currency pair from Ducascopy broker. Figure 5 depicts dependences between the probability of an increase (of one discretization unit) and the duration of a decrease (Figure 5a), and an increase wave (Figure 5b). Presented results suggest, that the shorter the duration of a wave, the higher the probability of change in a direction opposite to the direction of current wave. This relation is especially visible in case of decrease waves. This kind of effect can be caused by investors' different reaction to falls and rises [10]. Similar tendencies are also observed fir other parameters of WDABTR. Research results confirm that wave duration has a significant influence on the direction probability.

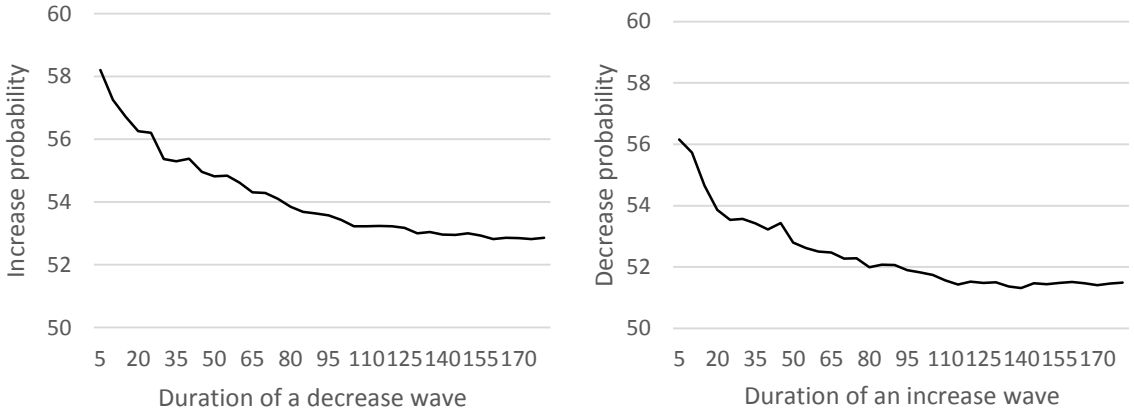


Figure 5. Dependences between increase probability and duration of an a) decrease and b) increase wave

We will now consider a relation between the registered wave length and the probability of future change direction. Figure 6 presents the relationship between the probability of an increase and the length of decrease wave (Figure 6a) and an increase wave (Figure 6b). With the growing length of a wave, the probability of an increase falls. Opposite relation occur in case of a decreasing wave. Identical tendencies are obtained for other WDABTR parameters.

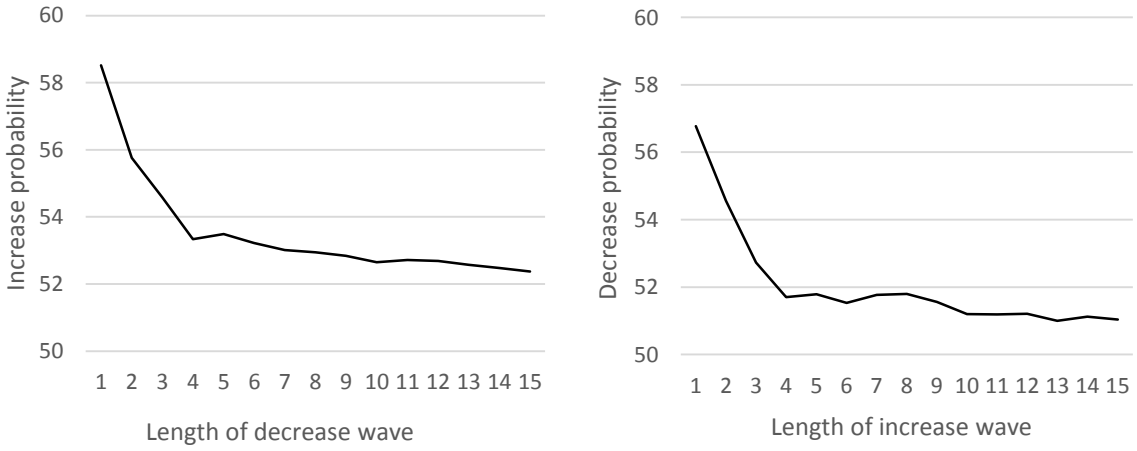


Figure 6. Dependences between increase probability and length of an a) decrease and b) increase wave

Research results suggest that the currency market has a wave structure. Both length and duration of a wave have an influence on the probability of future change direction. For waves short lasting less than 5 minutes, the probability of a future change in a direction opposite to the direction of current wave is the highest Identical tendency can be observed for other *m* and *p* parameter sets of WDABTR. About 58% value of future change direction assessment allows for a construction of HFT systems characterized by a positive return rate. With roughly 2-pips spread levels, the 53% prediction probability level allows for a stable profit.

5 Summary

In the article, a binary-temporal representation was used for modelling currency pair course trajectory. This kind of representation includes more information about the variability of a course than the candlestick one. Moreover, it can be used for modelling trajectories in HFT, i.e. the probability of a future change is equal to the probability of achieving a profit or loss in a single transaction. Next, a new WDABTR was proposed, which allows for wave detection in binary-temporal representation. The algorithm was used as a base in wave representation, which encompasses current information about waves occurring on the market and their parameters. Wave representation made it possible to perform a statistical verification of wave structure for given exchange rate.

Presented results, obtained based on 5-year historical data for ADU/NZD currency pair in binary-temporal representation with 15-pips discretization unit, confirmed the wave structure of the market. Research indicated an existence of visible dependences between wave parameters and future changes direction, especially for duration and length of a wave. Presented relations are characterized by an assessment of future change direction probabilities of a 58% prediction level. This value, regarding spreads offered on the market, allows for creating HFT systems with positive return rate.

References

[1] Aldridge I.: *High-frequency trading: a practical guide to algorithmic strategies and trading systems*, John Wiley and Sons, 2009.

[2] Bickford, J. L.: *Forex Wave Theory-A Technical Analysis for Spot and Futures Currency Traders*, The Mc-Graw-Hill Companies, New York, 2007.

[3] Burgess, G.: *Trading and Investing in the Forex Markets Using Chart Techniques* (Vol. 543), John Wiley and Sons, 2010.

[4] De Villiers, V.: *The Point and Figure Method of Anticipating Stock Price Movements: Complete Theory and Practice*, Stock Market Publications, 1933.

[5] Frost, A. J., and Prechter, R. R.: *Elliott wave principle: key to market behavior*. Elliott Wave International, 2005.

[6] Kirkpatrick, C. D., and Dahlquist, J. A.: *Technical analysis: the complete resource for financial market technicians*. FT press, 2010.

[7] Lo A. W., Mamaysky H., Wang J.: *Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation*. The Journal of Finance, 55(4), 1705-1770, 2000

[8] Logue, D. E., and Sweeney, R. J.: *White noise in imperfect markets: the case of the franc/dollar exchange rates*. The Journal of Finance, 32(3), 761-768, 1977.

[9] Murphy, J. J.: *Technical analysis of the futures markets: A comprehensive guide to trading methods and applications*, Prentice Hall Press, 1986.

[10] Oberlechner, T.: *The psychology of the foreign exchange market* (Vol. 506). John Wiley and Sons, 2005.

[11] Schlossberg, B.: *Technical Analysis of the Currency Market*. John Wiley & Sons Inc, 2006.

[12] Stasiak, M. D.: *Modelling of currency exchange rates using a binary representation,* Proceedings of 37th International Conference on Information Systems Architecture and Technology – ISAT 2016, Springer, 2016.

[13] Stasiak, M. D.: *Modelling of currency exchange rates using a binary-temporal representation*, Proceedings of: International Conference on Accounting Finance and Financial Institution: Theory meets Practice ICAFFI 2016, Springer, 2016. (accepted for publication).

Linguistic approximation of values close to the gain/loss threshold

Jan Stoklasa[1], Tomáš Talášek[2]

Abstract. Linguistic approximation (LA) is a natural last step of linguistic fuzzy modelling, providing linguistic labels (with their meaning known to the decision makers and understood well by them). Linguistic approximation techniques are based on approximation and hence the nature of the approximated output of mathematical model can be altered a bit by the application of these methods. LA can be considered beneficial in linguistic fuzzy modelling, as long as the interpretability and understandability of the provided linguistic outputs outweighs the possible loss/distortion of information. In many cases the distortion of information might be small and as such completely acceptable. Recently, however, Stoklasa and Talášek (2015) pointed out that when specific thresholds are of importance in the decision-making situation (e.g. the border between gains and losses), LA can distort the outcome of the decision-making situation by providing a loss label for a gain and vice-versa. In this paper, we investigate the phenomenon under different linguistic scales used for the approximation and provide a thorough discussion of this phenomenon in the context of linguistic approximation.

Keywords: Linguistic approximation, gains, losses, threshold, distance, linguistic scale.

JEL classification: D81, C44
AMS classification: 90B50, 91B06

1 Introduction

Mathematical models for economic practice and for managerial decision support (including e.g. investment decision support models, evaluation models) require a suitable interface to facilitate the exchange of information between the model and its users. Linguistic fuzzy modelling provides such an interface in terms of presenting the model and its outputs in terms of natural language [9]. To build a linguistic fuzzy model capable of providing understandable linguistic outputs to its users, we need to be able to transform the mathematical objects computed by the model into natural language. The process of transformation of the mathematical outputs of models into natural language is called linguistic approximation. There are various approaches to linguistic approximation (see e.g. [23] for an overview and [10, 15] for additional analysis of some of the methods). The majority of the methods of linguistic approximation is based on finding the fuzzy object (usually a fuzzy number) with a known linguistic label - e.g. methods finding the fuzzy set with a known linguistic label which is the closest (w.r.t. some distance measure, see e.g. [3] or) or the most similar (w.r.t. some similarity measure) to the approximated object. The performance of different similarity and distance measures has been recently studied in several papers (see e.g. [16, 17, 20]). Alternatively, there are also methods that use linguistic hedges and connectives to combine fuzzy sets with a known linguistic label to create and object close or similar enough to the approximated one (see e.g. [1, 5, 20, 22]). New methods for linguistic approximation are also being developed [19] and alternative uses for linguistic approximation are being considered (e.g. ordering of fuzzy numbers in [18], or conveying/stressing of specific pieces of information [23], the importance of the linguistic level of models has recently been discussed also in [2, 8, 9, 11, 12, 13, 14]). Clearly, linguistic modelling and linguistic approximation are topics that currently deserve the attention of researchers.

Although research into the behavioral aspects of linguistic approximation has already started, there are still several issues that need attention. For one the reliance on the distance or similarity measures in linguistic approximation to find the best fitting approximating linguistic term (in terms of the distance or similarity of its fuzzy-set-meaning to the approximated object) can prove problematic, since low distance and semantic closeness might not always be the same thing. Stoklasa and Talášek [10, p. 965, Figure 4] discuss the existence of a possible drawback of the use of linguistic approximation based on distance or similarity measures in the context of

[1]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic and Lappeenranta University of Technology, School of Business and Management, Skinnarilankatu 34, Lappeenranta, Finland, jan.stoklasa@upol.cz.

[2]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic, tomas.talasek@upol.cz.

economic decision-making concerning gains and losses. As already identified by Kahneman and Tversky e.g. in the context of prospect theory [6], the decision making and attitude to risk might be different based on the framing of a particular value as loss or gain by a specific decision maker. In this paper, we aim to investigate the possible drawbacks of the use of distance and similarity in linguistic approximation of the outputs of mathematical models in financial units, provide a closer-to-real-life example of the possible problems and analyze the performance of a frequently used distance measure and its possible alternative under different linguistic variables used for the linguistic approximation.

The paper therefore continues by a chapter summarizing the necessary theory and notation for linguistic fuzzy modelling including linguistic variables, linguistic scales and the basic idea of linguistic approximation. The next section specifies the problem under investigation, introduces the distance measures the performance of which will be investigated in this paper in the context of linguistic approximation of gains and losses and also specifies the linguistic variables that will be studied. A prototype example of the problem is also presented in this section. The next section summarizes the results of a numerical analysis of the performance of the selected methods and discusses the results and the last section draws conclusions for the paper.

2 Preliminaries

Let U be a nonempty set (the universe of discourse). A *fuzzy set* A on U is defined by the mapping $A : U \to [0, 1]$. For each $x \in U$ the value $A(x)$ is called the *membership degree* of the element x in the fuzzy set A and $A(.)$ is called the *membership function* of the fuzzy set A. $\text{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes the α*-cut* of A for any $\alpha \in [0, 1]$, $\text{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes the *support* of A. The *cardinality* of a fuzzy set A is computed as $\text{Card}(A) = \int_U A(x)dx$. A real-number characteristic representing the location of the fuzzy set A in the universe of discourse U is called the *center of gravity*: $\text{COG}(A) = \int_U xA(x)dx/\text{Card}(A)$.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: a) $\text{Ker}(A) \neq \emptyset$ (A is *normal*); b) A_α are closed intervals for all $\alpha \in (0, 1]$ (this implies A is *unimodal*); c) $\text{Supp}(A)$ is bounded. The family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on $[a, b]$, if $\text{Supp}(A)$ is a subset of an interval $[a, b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_2, a_3] = \text{Ker}(A)$ and $[a_1, a_4] = \text{Cl}(\text{Supp}(A))$, where $\text{Cl}(\text{Supp}(A))$ denotes a closure of $\text{Supp}(A)$. Each fuzzy number A can be also represented in the form of $A = \left\{[\underline{a}(\alpha), \overline{a}(\alpha)]\right\}_{\alpha \in [0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall \alpha \in (0, 1]$, and the closure of the support of A, $\text{Cl}(\text{Supp}(A)) = [\underline{a}(0), \overline{a}(0)]$. A fuzzy number A is called *linear* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$; for such fuzzy numbers we will use a simplified notation $A = (a_1, a_2, a_3, a_4)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by ordered triplet $A = (a_1, a_2, a_4)$. More details on fuzzy numbers and computations with them can be found for example in [4].

A *fuzzy scale* on $[a, b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on $[a, b]$, where for all $x \in [a, b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. A *linguistic variable* [24] is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is the name of the variable, $\mathcal{T}(\mathcal{V})$ is the set of its linguistic values (terms), X is the universe on which the meanings of the linguistic values are defined, G is a syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X. Linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ is called a *linguistic scale* on $[a, b]$ if $X = [a, b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$ and $\mathcal{M}(\mathcal{T}_i) = T_i, i = 1, \ldots, s$ form a fuzzy scale on $[a, b]$.

3 Definition of the problem - distance based linguistic approximation in the gain/loss domain

Kahneman and Tversky (see e.g. [7, 21]) suggested and subsequently experimentally proved, that the carrier of decision-power in real life situations concerning e.g. sums of money is not the absolute value, but its reframing into gain or loss. They also postulate, that people deal differently with gains and losses (willingness to take risk might change, see [6] for more). The purpose of linguistic approximation is to find the best linguistic label for a given mathematical output. If we assume a fuzzy number O to be linguistically approximated by one of the linguistic values of a linguistic scale $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$, then a distance based approach to linguistic approximation translates into (1), i.e. into finding such an element in $T(V)$, for which the distance d of its fuzzy-number meaning to the approximated fuzzy output O is minimal.

$$T_O = \arg \min_{\mathcal{T}_i \in \mathcal{T}(\mathcal{V})} d(T_i, O) \tag{1}$$

We need to stress here, that the linguistic approximation is not always able to preserve all the information carried out by the approximated output (hence "approximation"). We, however, need to make sure, that the most important characteristics of the approximated objects are not distorted too much. In the context of gains/losses, we would at least expect a clear loss not to be assigned a "gain" label and vise-versa. The outcome of the linguistic approximation obviously depends on the linguistic variable used in the process and on the definition of the meaning of its linguistic values. In this paper, we assume two different general types of linguistic scales for the purpose (the meanings of the linguistic values of both of them are summarized in Figure 1). The first linguistic scale assumes a decision maker not distinguishing in the loss domain, while the other one assumes that losses and gains are partitioned in a similar manner, the red lines represent the loss/gain threshold.

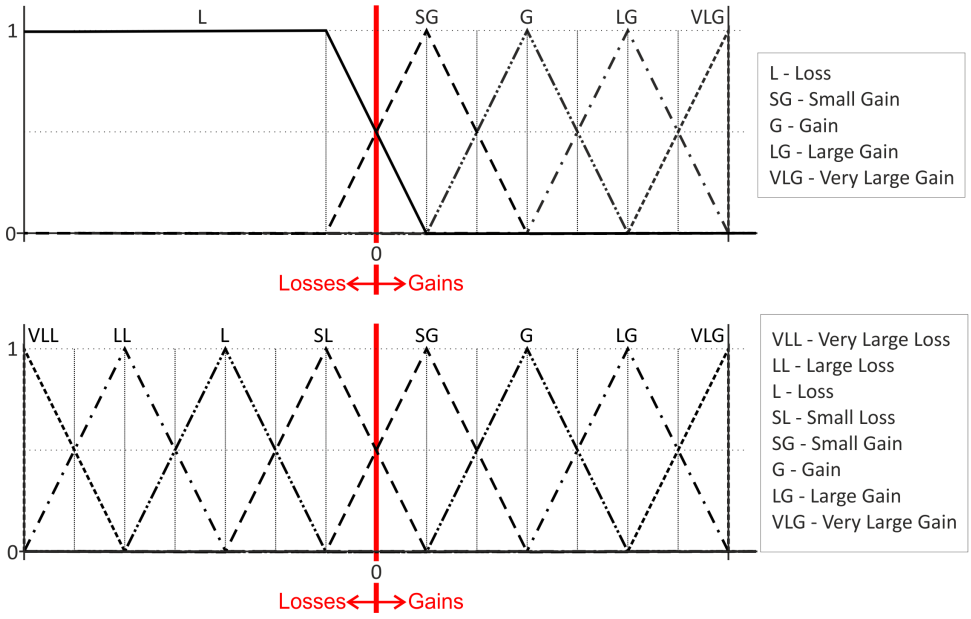


Figure 1: The scales used for linguistic approximation of the outputs of mathematical models representing financial values (e.g. NPV of a project, etc.) or future cash flow estimates. The case represented in the top figure does not differentiate in the area of losses, the bottom linguistic scale differentiates in the area of losses in the same way as in gains.

Obviously the other crucial factor influencing the outcome of the approximation is the distance measure used. One of the frequently used distance measures of fuzzy numbers is the *dissemblance index* of fuzzy numbers A and B, $d_1(A,B)$, defined by the formula (2). The dissemblance index requires both A and B to be fuzzy numbers, which is not a problem, since the meanings of the linguistic values of the approximating linguistic variable are usually represented by fuzzy numbers and the approximated object can be expected to be a fuzzy number as well. Without any loss of generality we use the dissemblance index in a non-normalized form, if needed, it can be normalized so that its value lies within the $[0,1]$ interval, i.e. $d_1(A,B)/2\,(b-a) \in [0,1]$ for $A,B \in \mathcal{F}_N([a,b])$. Note, that in the gain/loss domain, we are expecting the outputs of the mathematical models to be fuzzy quantities (e.g. represented by triangular fuzzy numbers).

$$d_1(A,B) = \int_0^1 |\underline{a}(\alpha) - \underline{b}(\alpha)| + |\overline{a}(\alpha) - \overline{b}(\alpha)|\, d\alpha, \tag{2}$$

Using d_1 and the top linguistic scale in Figure 1, we can obtain very counterintuitive results of linguistic approximations. An example of such a problematic result is presented in Figure 2, where a clear "loss" represented by the fuzzy number Out is linguistically approximated by the label "small gain". Such a mislabelling of an output can have serious consequences in decision support, since a "gain" label can motivate a different reaction of the decision maker than would be required for an actual loss Out. Note, that in Figure 2, we have $d_1(Out,SG) = a+b < \int_0^1 |\underline{l}(\alpha) - \underline{Out}(\alpha)|d\alpha < d_1(L,Out) = \int_0^1 |\underline{l}(\alpha) - \underline{Out}(\alpha)| + |\overline{l}(\alpha) - \overline{Out}(\alpha)|d\alpha$.

We have thus identified an even clearer example of the possible problem with distance-based linguistic approximation, where even though a mathematically sound distance measure and a reasonable linguistic scale is used, the resulting approximation can completely change the nature of (information carried by) the actual approximated output. In the next section, we investigate how serious this problem is for the dissemblance index and compare the performance of this distance measure with another distance measure - namely the modified Bhattacharyya distance

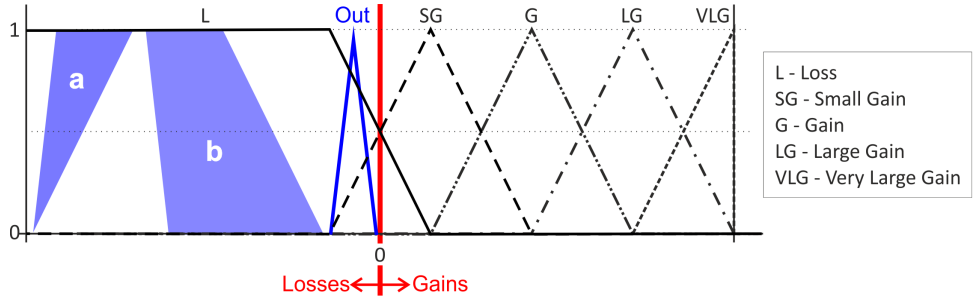


Figure 2: A graphical representation of $d_1(Out, SG)$ represented by the blue shapes a) and b). Note, that $a = \int_0^1 |\underline{Out}(\alpha) - \underline{sg}(\alpha)|d\alpha$, $b = \int_0^1 |\overline{Out}(\alpha) - \overline{sg}(\alpha)|d\alpha$ and $a + b = d_1(Out, SG)$. Clearly $d_1(Out, SG) < d_1(Out, L)$ and thus "small gain" is considered to be a better linguistic approximation for Out than "loss", even though Out is completely in the loss domain.

of fuzzy numbers d_2, which can be computed in the following way:

$$d_2(A, B) = \left[1 - \int_U (A^*(x) \cdot B^*(x))^{1/2} dx\right]^{1/2}, \tag{3}$$

where $A^*(x) = A(x)/\text{Card}(A)$ and $B^*(x) = B(x)/\text{Card}(B)$. We also investigate how a change in the linguistic scale used for the approximation influences the results of the approximation and the performance (and appropriateness) of both distance measures.

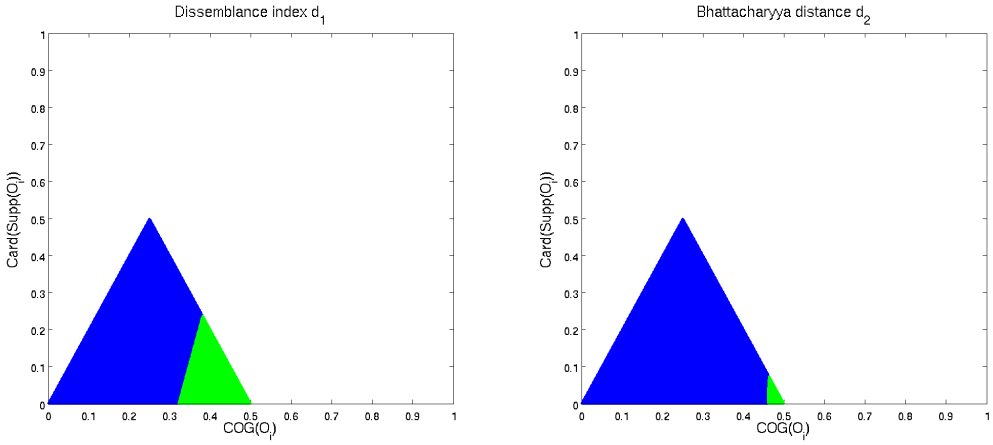


Figure 3: Results of the numerical experiment for the linguistic scale not differentiating in the loss domain. Each point represents one symmetrical triangular fuzzy number O_i, $\text{Supp}(O_i) \subseteq [-r, 0], i = 1, \ldots, 125\,000$, characterized by its center of gravity (x-coordinate, the $[-r, r]$ universe is just linearly transformed to $[0, 1]$) and the cardinality of its support (y-coordinate). The colour represents the result of the linguistic approximation: blue for *loss* and green for *small gain*. Results are presented for the linguistic approximation using d_1 (left plot, 103 758 fuzzy numbers approximated correctly as losses, 21 242 incorrectly as gains) and d_2 (right plot, 123 484 fuzzy numbers approximated correctly as losses, 1 516 incorrectly as gains).

4 Numerical analysis and discussion of the results

To stress the magnitude of the problem of possible mislabelling of "losses" by a "gain" label, we will consider only fuzzy-number outputs of the mathematical model to be approximated which are completely in the domain of losses, i.e. for which the whole support lies in the loss domain. To simplify the analysis, we will also assume the approximated objects are symmetrical triangular fuzzy numbers. Using the same approach as in [17], a total of 125 000 symmetrical triangular fuzzy numbers O_i were generated, $i = 1, \ldots, 125\,000$, which uniformly cover the $[-r; 0]$ universe, where r represents the maximum expected gain and $-r$ the maximum expected loss. The results of the numerical experiment using the top linguistic scale from Figure 1 are presented in Figure 3. Almost 17% of the triangular fuzzy numbers representing a clear loss are mislabelled as gains using the dissemblance index. Note also, that low-uncertain fuzzy numbers can still be labelled as gains, even though their COG is close to the middle of $[-r; 0]$ interval. The dissemblance index clearly is not a good choice with a linguistic scale which treats

gains and losses asymmetrically, since the large cardinality of the L fuzzy number distorts the computations. On the other hand the use of Bhattacharyya distance in this case can significantly reduce the risk of mislabeling - see that only about 1.5% of the clear loss fuzzy outputs were labelled as gains - all of them with low cardinality and COG close to the loss/gain threshold.

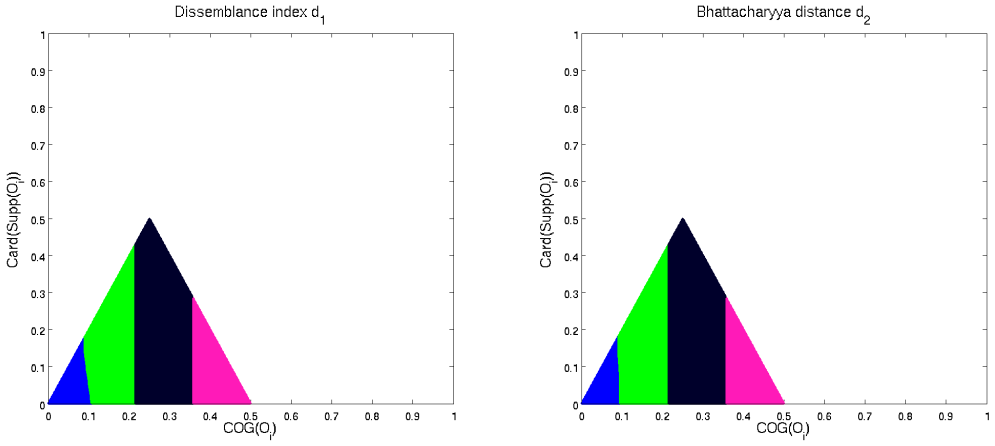


Figure 4: Results of the numerical experiment for the linguistic scale differentiating in the loss domain in the same way as in the gains domain. Each point represents one symmetrical triangular fuzzy number O_i, $\mathrm{Supp}(O_i) \subseteq [-r, 0], i = 1, \ldots, 125\,000$, characterized by its center of gravity (x-coordinate, the $[-r, r]$ universe is just linearly transformed to [0,1]) and the cardinality of its support (y-coordinate). The colour represents the result of the linguistic approximation: blue for *very large loss*, green for *large loss*, black for *loss* and purple for *small loss*. Results are presented for the linguistic approximation using d_1 (left plot, 12 620 times VLL, 49 616 times LL, 55 278 times L , 6 972 times SL and in 514 cases two loss labels were suggested with the same distance to O_i) and d_2 (right plot, 11 862 times VLL, 50 387 times LL, 55 278 times L , 6 972 times SL and in 501 cases two loss labels were suggested with the same distance to O_i).

The same analysis was also performed for the symmetrical linguistic scale presented in the bottom part of Figure 1. The results are presented in Figure 4. For a symmetrical underlying linguistic scale the differences between the distance measures are almost nonexistent. Also note, that a gain label was never assigned for a symmetrical triangular fuzzy number representing a clear loss.

We can clearly see that both the selection of the linguistic scale and the selection of the distance method can significantly influence the results of the linguistic approximation. As Bhattacharyya distance favours supersets, it seems to be a method of choice for the use with linguistic scales which are not symmetrical with respect to the loss/gain threshold. On the other hand the selection of a symmetrical linguistic scale can get rid of the mislabelling problem between gains and losses and renders the performance of the two investigated distance measures almost identical.

5 Conclusion

This paper investigates the performance of two different distance measures of fuzzy numbers in the distance-based linguistic approximation of fuzzy numbers representing uncertain sums of money in the gain/loss framing. It provides a clear example of possible mislabeling problem, where as a result of the choice of a selection of an improper distance measure losses can be linguistically labelled as gains (and by the same logic gains as losses). In the context of the findings of prospect theory, this presents a significant problem in decision support, since gains and losses can motivate different decision strategies. A numerical analysis of this problem is performed and two possible solutions of the problem - the use of symmetrical linguistic scales or the use of Bhattacharyya distance method are suggested. The paper presents a first step in the investigation of the performance of linguistic approximation methods in the gain/loss domain, the investigation of the role of other distance and similarity measures as well as the implications of different formats of linguistic scales will be the natural next steps of this research stream.

Acknowledgements

Supported by the grant No. IGA_FF_2017_011 of the Internal Grant Agency of Palacký University Olomouc.

References

[1] Bonissone, P. P.: A fuzzy sets based linguistic approach: theory and applications. In: *Proceedings of the 12th conference on Winter simulation*, 1980. 99–111.

[2] Collan, M., Stoklasa, J., and Talašová, J.: On Academic Faculty Evaluation Systems: More than just Simple Benchmarking. *International Journal of Process Management and Benchmarking* **4** (2014), 437–455.

[3] Degani, R., and Bortolan, G.: The problem of linguistic approximation in clinical decision making. *International Journal of Approximate Reasoning* **2** (1988), 143–162.

[4] Dubois, D., and Prade, H., eds.: *Fundamentals of Fuzzy Sets*. Kluwer Academic Publishers, Massachusetts, 2000.

[5] Eshragh, F., and Mamdani, E. H.: A general approach to linguistic approximation. *International Journal of Man-Machine Studies* **11** (1979), 501–519.

[6] Kahneman, D., and Tversky, A.: Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47** (1979), 263–292.

[7] Kahneman, D., and Tversky, A.: Choices , Values , and Frames. *American Psychologist* **39** (1984), 341–350.

[8] Krejčí, J., and Stoklasa, J.: Fuzzified AHP in the evaluation of scientific monographs. *Central European Journal of Operations Research* **24** (2016), 353–370.

[9] Stoklasa, J.: *Linguistic models for decision support*. Lappeenranta University of Technology, Lappeenranta, 2014.

[10] Stoklasa, J., and Talášek, T.: Jazykově orientované modelování pro ekonomickou a manažerskou praxi některé otevené problémy. In: *Znalosti pro tržní praxi 2015*. 959–969.

[11] Stoklasa, J., and Talášek, T.: On the use of linguistic labels in AHP: calibration, consistency and related issues. In: *Proceedings of the 34th International Conference on Mathematical Methods in Economics*. Technical University of Liberec, Liberec, 785–790.

[12] Stoklasa, J., Talášek, T., Kubátová, J., and Seitlová, K.: Likert scales in group multiple-criteria evaluation. *Journal of Multiple-Valued Logic and Soft Computing* (2017), (in press).

[13] Stoklasa, J., Talášek, T., and Musilová, J.: Fuzzy approach - a new chapter in the methodology of psychology? *Human Affairs* **24** (2014), 189–203.

[14] Stoklasa, J., Talášek, T., and Stoklasová, J.: Semantic differential and linguistic approximation - identification of a possible common ground for research in social sciences. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016*. Societas Scientiarum Olomucensis II, Olomouc, 495–501.

[15] Talášek, T., and Stoklasa, J.: Jazykově orientované modelování pro ekonomickou a manažerskou praxi - vybrané metody pro jazykovou aproximaci. In: *Znalosti pro tržní praxi 2015*. 977–989.

[16] Talášek, T., and Stoklasa, J.: The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016*. Societas Scientiarum Olomucensis II, Olomouc, 539–546.

[17] Talášek, T., and Stoklasa, J.: A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation Under Different Linguistic Scales. *Journal of Multiple-Valued Logic and Soft Computing* (2017), (in press).

[18] Talášek, T., Stoklasa, J., Collan, M., and Luukka, P.: Ordering of fuzzy numbers through linguistic approximation based on Bonissone's two step method. In: *16th IEEE International Symposium on Computational Intelligence and Informatics*. 285–290.

[19] Talášek, T., Stoklasa, J., and Talašová, J.: Linguistic approximation using fuzzy 2-tuples in investment decision making. In: *Proceedings of the 33rd International Conference on Mathematical Methods in Economics*. 817–822.

[20] Talášek, T., Stoklasa, J., and Talašová, J.: The role of distance and similarity in Bonissone's linguistic approximation method a numerical study. In: *Proceedings of the 34th International Conference on Mathematical Methods in Economics*. Technical University of Liberec, Liberec, 845–850.

[21] Tversky, A., and Kahneman, D.: Rational Choice and the Framing of Decisions. *The Journal of Business* **59** (1986), S251–S278.

[22] Wenstøp, F.: Quantitative analysis with linguistic values. *Fuzzy Sets and Systems* **4** (1980), 99–115.

[23] Yager, R. R.: On the retranslation process in Zadeh's paradigm of computing with words. *IEEE transactions on systems, man, and cybernetics. Part B: Cybernetics* **34** (2004), 1184–1195.

[24] Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences* **8** (1975), 199–249.

Macroeconomic modelling using cointegration vector autoregression

Radmila Stoklasová[1]

Abstract. The article deals with modern econometric methods that will be used in long-term structural macro econometric modeling of the Czech economy. The structural model is estimated for the Czech economy to quarterly data 2005Q1 - 2016Q4. On the basis of the economic theory is to derive long-term relationships for a small open economy with 9 endogenous variables (6 domestic and 3 international) and 1 exogenous variable. The data source was the Eurostat database, FRED, Czech National Bank and the Czech Statistical Office.
This article aims to find cointegration equations for modeling the long-term equilibrium of economic relations in the Czech Republic in the analyzed period. Based on the number of determined cointegration relationships is tested weak exogenous and there are tested hypotheses considering by restrictions on the coefficients so that the estimated cointegration relations in accordance with the economic and statistical theories. Achieved empirical results are influenced by the fact that the Czech economy has undergone in the period currency crisis. The calculations used EViews software version 9.

Keywords: ADF test of stationarity, correlation analysis, Granger causality, time series analysis, VECM model.

JEL Classification: C10, E27
AMS Classification: 62P20, 91B64

1 Derivation of the basic model

Cointegration vector autoregression (CVAR) models pay attention to the dynamics of economic variables because economic theories offer conclusions regarding cointegration of time series. The main idea is that the economic variables deviate from their equilibrium values in the short run, while being drawn to these equilibriums in the long run. Forces shifting equilibrium states are applied in the long run, leading to stochastic development trends, but the forces drawing variables to equilibriums are applied as well which lead to the cointegrating relationships between them. As stated by Hoover [6], the macroeconomic time series are characterized by unsteadiness and cointegratability.

CVAR models began to emerge at the turn of 1980s and 1990s and were focused on the US economy, as stated in the study [2] and [9]. Garrat et al. [4] created a model of the economy of the United Kingdom in their study. This model contains five equilibrium relationships. These relationships were derived from economic theory and, therefore, have a structural character. The model includes purchasing power parity in relative version (PPP), money demand (MD), a gap between domestic and foreign product (GR), interest rate parity (IRP), Fisher inflation parity (FIP). A cointegration analysis showed that long-run structural equilibrium relationships correspond to empirical cointegration relationships, so the model used is suitable for a small open economy. A similar methodology was used in the case of the German economy, as stated in Schneider et al. in [11]. The authors tested the same model on the German quarterly data for the period from 1991 to 2005. VAR(2) model was also used in this work, and the presence of five cointegration relationships was discovered. Paper [1] is focused on the Swiss economy in the period from Q1/1976 to Q4/2006. Cointegration tests in this work indicate the presence of five cointegration relationships as well. Paper [6] is focused on the Czech economy in the period from Q1/1995 to Q4/2009 and it also confirms the presence of five cointegration relationships.

1.1 Description of data files

Based on the economic theory as in [10] and economic derivation of relationships (1)–(4) for a small open economy, we use 9 endogenous variables, of which 6 domestic and 3 foreign, and an exogenous variable of oil price, which represents a long-run forcing variable, for estimating the VAR(p) model.

[1] School of Business Administration in Karviná, Silesian University in Opava,
Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech Republic,
e-mail: stoklasova@opf.slu.cz.

Labeling and meaning of variables:

- $PP_t = \ln(P_t)$, where P_t is the domestic price index of industrial producers (2005 = 100),
- $PE_t = \ln(P1_t)$,where $P1_t$ is index total manufacturing production for the Eurozone (2005 =100),
- $POI_t = \ln(P3_t)$, where $P3_t$ is the oil price index (2005 = 100),
- $Y_t = \ln(YC_t / P_t)$,where YC_t is the domestic nominal GDP (CZK millions),
- $YE_t = \ln(Y1_t / P1_t)$, where $Y1_t$ is foreign nominal GDP in Eurozone (EUR millions),
- $E_t = \ln(E1_t)$, where $E1$ is the nominal exchange rate index CZK/EUR (2005 = 100),
- $R_t = 0.25\ln(1+ (R1_t/100))$, where $R1_t$ is the annual domestic interest rate (PRIBOR_1Y),
- $RE_t = 0.25\ln(1+(R2_t/100))$, where $R2_t$ is the annual foreign interest rate (EURIBOR_1Y),
- $M2 = \ln(M\_2_t)$, where M_2_t is monetary aggregate $M2$ in mil. CZK,
- $INF_t = PS_t - PS_{t-1}$.

Five equations representing the long-run equilibrium paths of macroeconomic development:

$$PPP: \quad PP_t - PE_t - E_t = b_1 + \varepsilon_{1t} \tag{1}$$

$$MD: \quad M2_t - Y_t = b_2 + \beta_{21}R_t + \beta_{22}Y_t + \varepsilon_{2t} \tag{2}$$

$$GR: \quad Y_t - YE_t = b_3 + \varepsilon_{3t} \tag{3}$$

$$IRP: \quad R_t - RE_t = b_4 + \varepsilon_{4t} \tag{4}$$

$$FIP: \quad R_t - INF_t = b_5 + \varepsilon_{5t} \tag{5}$$

Domestic and foreign product

Both variables do not show the same trend in the period from 2005 to 2008, it is a period when there was a mortgage crisis in 2007 that expanded to a financial crisis. Governments and central banks implemented measures to support the financial sector since September 2008. The time series of foreign products shows a drop to 9.86 in Q3/2008 and a growth to 9.95 in Q4/2008. The financial crisis became a crisis of the real economy in early 2009. Both variables show a downward trend in the period from 2009 to 2012. Both variables show a growing trend in the period from 2013 to 2016, leading to economic growth.

2 Macroeconomic model for the Czech economy

The second chapter deals with the estimation and testing of the basic macroeconomic model for the Czech economy in the period from Q1/2005 to Q4/2016. The chapter is divided into five parts. First, a stationarity of VAR model variables is tested. The second step is to determine the order of VAR model and its diagnostics. This subchapter includes the testing of weak exogeneity of oil price. As the third step, the cointegration relationships for the VAR(*p*) model are tested using the Johansen's method and a number of cointegration relationships is determined. The fourth step is to estimate a VECM(*p–1*) model assuming the existence of *r* cointegration relationships. The last subchapter provides equations for long-run equilibrium relationships.

2.1 Stationarity testing

The preparatory phase of estimating the VAR model is testing the stationarity of variables included in the model or their first differences. The test results for all variables are provided in Table 1. Dickey-Fuller test (ADF) was used to test the stationarity. The second column provides information on the model type of testing unit root (n = no trend and level constants /c = constant /c+t = level constant and trend), the third column contains the calculated T-statistics; the following column contains the corresponding level of statistical significance. The last column includes the result of testing: N = non-stationary, (H0 not rejected), S = stationary (H0 rejected).

variable	n/c/c+t	T-stat	significance	result	variable	n/c/c+t	T-stat	significance	result
PP_t	c+t	-0.698	0.967	N	$D(PP_t)$	n	-5.466	0.000	S
PE_t	c+t	-1.538	0.802	N	$D(PE_t)$	n	-5.812	0.000	S
POI_t	c+t	-2.381	0.384	N	$D(POI_t)$	n	-6.592	0.000	S
Y_t	c+t	-3.076	0.125	N	$D(Y_t)$	c	-4.754	0.000	S
YE_t	c+t	-0.899	0.947	N	$D(YE_t)$	n	-6.334	0.000	S
E_t	c+t	-3.056	0.127	N	$D(E_t)$	n	-3.456	0.054	N
R_t	c+t	-1.043	0.263	N	$D(R_t)$	n	-4.191	0.001	S
RE_t	c+t	-2.165	0.497	N	$D(RE_t)$	n	-5.507	0.000	S

$M2$	c+t	-2.623	0.272	N	$D(M2)$	c	-9.739	0.000	S
INF_t	n	-0.879	0.328	N	$D(INF_t)$	n	-3.264	0.062	N

Table 1 Testing the unit root of the variables in levels and their first differences

The variables (PP_t; PE_t; POI_t; Y_t; YE_t; R_t; RE_t; M) for VAR model exhibit the properties of first-order non-stationarity, i.e. I(1); therefore, the long-run cointegration relationships may exist between these variables.

2.2 Estimation and diagnostics of VAR model

This chapter describes the estimation of VAR(p) model. Model estimation includes a determination of p order for delayed variables (PP_t; PE_t; POI_t; Y_t; YE_t; R_t; RE_t; M) in a vector autoregressive model. This delay level is usually the same for all VAR model equations. Given the set of quarterly data, delay time of 4 maximum was considered. Table 2 summarizes the results based on the minimization of selected criteria: FPE: Final prediction error, AIC: Akaike information criterion, SC: Schwarz information criterion, HQ: Hannan-Quinn information criterion a LR: likelihood ratio, which is based on the principle of maximum likelihood.

Lag	LogL	LR	FPE	AIC	SC	HQ
0	1008.444	NA	2.46e-30	-45.47474	-45.15034	-45.35443
1	1387.422	602.9188	1.56e-36	-59.79190	-56.87231	-58.70917
2	1489.090	124.7745*	3.77e-37	-61.50408	-55.98931	-59.45894
3	1585.294	83.08560	2.14e-37	-62.96792	-54.85796	-59.96036
4	1751.195	82.95063	2.12e-38*	-67.59979*	-56.89466*	-63.62981*

* indicates lag order selected by the criterion

Table 2 VAR lag order selection criteria

Most of these tests recommend a delay order of 4 except for LR criterion which recommends the VAR(2) model. Furthermore, Wald test was performed to test up to the delay of p = 1,2,3,4 for each equation. The results confirmed the choice of VAR(4) model. VAR(3) model was chosen for further considerations because the small number of observations at p = 4 would lead to a lower the explanatory power of estimated model.

The article also examines the weak exogeneity of POI_t oil price variable. The results of Granger causality test are provided by Table 3. The oil price variable is dependent on the domestic GDP in the VAR model and also on all the endogenous variables in the VAR(3) model at the 5% significance level; therefore, it cannot be considered a weak exogenous variable. The oil price variable will be excluded from the model as a forcing variable.

Excluded	Chi-sq	df	Prob.
M2	4.643332	3	0.2178
PET	0.566772	3	0.9040
PPT	5.353063	3	0.1477
RET	1.635223	3	0.6514
RT	1.479856	3	0.6869
YET	3.200838	3	0.3617
YT	*22.01445*	*3*	*0.0001*
ALL	*110.1927*	*3*	*0.0000*

Table 3 Testing the weak oil price exogeneity

The Czech economy model result is the VAR(3) model for endogenous variables (PP_t; PE_t; Y_t ; YE_t ; R_t; RE_t; M). This model without restrictions was estimated using the maximum likelihood ML method, as in [5].

The diagnostic test was performed to assess the quality of the VAR (3). These tests include the conditions of stationarity, non-correlatability and normality test of residual component. VECM(3) model stationarity conditions were verified using the graphical representation the inverse values of estimated autoregressive polynomial roots. These values lie within a unit circle, i.e. the VECM(3) model is stationary. Non-correlatability of residual component of the estimated VECM(3) model was tested using LM test. This test confirms the non-correlatability of residual component (at the 5% significance level, a null hypothesis of non-correlatability of residual component is not rejected). A residual component normality test was done using the Jarque-Bera test. The residual component normality null hypothesis was not rejected at the 5% significance level.

2.3 Cointegration relationship testing

This chapter deals with the testing of the number of cointegration relationships in VAR(3) model for the endogenous variables (PP_t; PE_t; Y_t; YE_t; R_t; RE_t; M) using the Johansen's method, as shown in [3], [8]. Table 4 confirms the existence of 5 cointegration relationships for VECM(2). This is a model that includes unlimited level constant and restricted trend component.

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None *	0.834555	269.5137	150.5585	0.0000
At most 1 *	0.702056	188.5534	117.7082	0.0000
At most 2 *	0.590460	134.0652	88.80380	0.0000
At most 3 *	0.576250	93.89277	63.87610	0.0000
At most 4 *	0.500881	55.25521	42.91525	0.0019
At most 5	0.288627	23.98425	25.87211	0.0843

Trace test indicates 5 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level

Table 4 Unrestricted Cointegration Rank Test (Trace)

2.4 Cointegration VECM and restrictions

The next step of the estimation procedure is to estimate VECM(p – 1) model, therefore VECM(2), assuming 5 cointegration relationships.

VECM(2) model contains restrictions listed in the Table 5. There are 28 restrictions in the matrix and these restriction: $\beta_1 + \beta_2 = 0$; $\beta_5 + \beta_6 = 0$.

	M2	PET	PPT	RET	RT	YET	YT
PPT	0	– 1	1	0	0	β_1	β_2
MD	1	0	0	0	β_3	0	β_4
GR	0	0	0	β_5	β_6	β_7	1
IRP	0	0	0	– 1	1	0	0
FIP	0	0	– 1	0	1	0	0

Table 5 Restrictions

VECM(2) model stationarity conditions are shown in Figure 1 which shows the inverse values of estimated autoregressive polynomial roots. These values lie within a unit circle, i.e. the VECM(2) model is stationary.

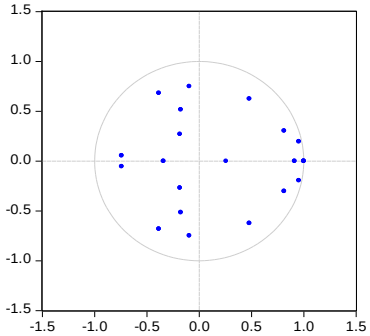


Figure 1 Inverse roots of AR characteristic polynomial of model VECM(2)

Non-correlatability of residual component of the estimated VECM(2) model was tested using LM test. Table 6 shows the test values. This test confirms the non-correlatability of residual component (at the 5% significance level, a null hypothesis of non-correlatability of residual component is not rejected).

Lags	LM-Stat	Prob.
1	46.55831	0.5727
2	60.18466	0.1314
3	63.11026	0.0848
4	57.93207	0.1790

Table 6 VECM(2) residual serial correlation LM tests

A residual component normality test was done using the Jarque-Bera test. The test results are in Table 7. The residual component normality null hypothesis was not rejected at the 5% significance level.

Component	Jarque-Bera	df	Prob.
1	5.501451	2	0.0639
2	5.557621	2	0.0621
3	0.288532	2	0.8657
4	1.312755	2	0.5187
5	0.730902	2	0.6939
6	2.322275	2	0.3131
7	0.097435	2	0.9525
Joint	15.81097	14	0.3251

Table 7 VECM(2) residual normality test

The residual component homoscedasticity null hypothesis was not rejected at the 5% significance level, as the results show Chi-sq = 1084.706, df = 1064; Prob. = 0.3226. The test "No Cross Terms" was performed (only levels and squares).

Table 8 shows the predictive ability of each estimated equation in VECM(2) with restrictions. The average level of adj.R^2 = 0.628 and it ranges within the interval from 0.427 (domestic economic output equation) to 0.755 (foreign economic output equation). These explanation measures compare favourably with the English [4] and Swiss study [1].

	D(M2)	D(PET)	D(PPT)	D(RET)	D(RT)	D(YET)	D(YT)
R^2	0.795	0.858	0.766	0.706	0.861	0.861	0.675
Adj.R^2	0.641	0.751	0.588	0.484	0.754	0.755	0.427

Table 8 Determination coefficients for the estimated equations of the resulting model

Finally, we can say that the estimation of the VECM(2) model with restrictions is stable, with a relatively high explanatory power. The residual component is not correlated; residual component heteroscedasticity and residual component non-normality were not demonstrated.

2.5 Equations of long-run relationships

Equations (6)–(10) summarize the results of estimating cointegration VECM(2) with 28 restrictions (stated in Table 5) for modelling the long-run equilibrium relationships for the Czech Republic in the researched period. All regression coefficients are statistically significant at the 5% significance level.

$$PPP: \quad PP_t - PE_t = 0.953 + 1.174(Y_t - YE_t) + \varepsilon_{1t} \tag{6}$$

$$MD: \quad M2_t = 10.818 - 6.668R_t + 0.418Y_t + \varepsilon_{2t} \tag{7}$$

$$GR: \quad Y_t = -2.63 - 17.743(R_t - RE_t) + 1.173YE_t + \varepsilon_{3t} \tag{8}$$

$$IRP: \quad R_t - RE_t = 0.00354 + \varepsilon_{4t} \tag{9}$$

$$FIP: \quad R_t - PP_t = -4.706 + \varepsilon_{5t} \tag{10}$$

3 Conclusion

This article dealt with the modelling of long-run structural equilibrium relationships for the Czech economy in the period from 2005 to 2016. Modelling using cointegration vector autoregressive model with restrictions for endogenous variables (PP_t; PE_t; Y_t; YE_t; R_t; RE_t; M) was used. Estimation of the VECM(2) model with restrictions is stable, with a relatively high explanatory power. This model includes five cointegration relationships with 28 restrictions. The residual component is not correlated; residual component heteroscedasticity and residual component non-normality were not demonstrated.

The first PPP equation (6) shows the considered effect of international product gap (coefficient 1.17%) on purchasing power parity (PPP). Just as in related foreign studies (British, German), the hypothesis on the validity of the relative PPP version in the long run cannot be rejected in this case. The second equation MD (7) shows that the demand for monetary balances is negatively dependent on domestic nominal interest rate (coefficient 6.67%) and positively dependent on real domestic product (coefficient 0.42%) which corresponds to monetary theory. Monetary aggregate M2 is large enough to cope with any changes in the tools used for financial intermediation. The third equation GR (8) shows that the real domestic product is affected by interest difference (R_t - RE_t) (coefficient 17.74%) and foreign product (coefficient 17.1%). A one percent increase (R_t - RE_t) leads to an almost 18-times decrease in the domestic product. Influence (R_t - RE_t) on the domestic product is consistent with theory. British and German authors came to the same statement. The last two equations IRP (9) and FIP (10) are in line with the values found for developed countries with a long tradition of market economy. Regarding the parity of domestic and foreign interest rates, British authors estimate the size of constant at 0.0058 and German authors at 0.002.

The model used for a small open economy is theoretically and empirically consistent because the estimated parameters have reasonable signs and values. Empirical results are influenced by the fact that the Czech economy underwent a currency crisis characterized by atypical behaviour of interest rates, monetary indicators, exchange rate and other indicators in the researched period. The currency crisis also affected the interaction between examined variables. It would be interesting to make an analysis which will examine and compare the situation before and after the crisis.

Acknowledgements

This paper was supported by the Ministry of Education, Youth and Sports Czech Republic within the Institutional Support for Long-term Development of a Research Organization in 2017.

References

[1] Assenmacher-Weschke,K., and Pesaran,M.H.: A VECX Model of the Swiss Economy. *CESifo Working Paper*,No. 2281. University of Cambridge, Cambridge, 2008.
[2] Blanchard, O. J., and Quah, D.: The Dynamic Effects of Aggregate Demand and Supply Disturbanes. *American Economic Review* **79**(4), (1989), 655-673.
[3] Enders, W.: Cointegration and Threshold Adjustment. *Journal of Business & Economic Statistics* **19**(2), (2001), 166-176.
[4] Garratt, A., and Lee,K.: A Long Run Structural Macroeconometric Model of the UK. (Pesaran, M.H., and Shin, Y., eds.). *The Economic Journal* **113**(487), (2003), 412-455.
[5] Hančlová, J.: Kvartální makroekonomický model české ekonomiky. In: *Aktuální aspekty české a světové ekonomiky*, Technická univerzita Liberec, Liberec, 2009, 138-147.
[6] Hančlová, J., and Kubicová, I.: *Makroekonometreické modelování české ekonomiky a vybraných ekonomik EU*. (Macháček,M.,Melecký, L.,Melecký,M.,Nevima,J. and Ramík,J.,eds.).VŠB-TU,Ostrava, 2010, 31 – 88.
[7] Hoover,K.D., and Johansen,S.:Allowing the Data to Speak Freely: The Macroeconometrics of the Cointegrated Vector Autoregression. (Juselius,K., eds.).*American Economic Review:Paper and Proceedings* **98**(2), (2008), 251-255.
[8] Johansen, S.: *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 1995.
[9] King, R.G., and Plosser, CH.I.: Stochastic Trends and Economic Fluctuations. (Stock, J.H., and Watson, M.W., eds). *American Economic Review* **81**(4), (1991), 819-840.
[10] Obstfeld, M., and Rogoff,K.: *Foundations of International Macroeconomics.* MIT Press, Cambridge, 1996.
[11] Schneider, E., and Chen,P.: A Long-Run Structural Macroeconometric Model for Germany. (Frohn,J., eds.). *Economics Discussion Papers: The Open-Access, Open-Assessment E-Journal*,No.47, 2007.

Pseudomedian in robustification of Jarque-Bera test of normality

Luboš Střelec[1], Ladislava Issever Grochová[2]

Abstract. The assumption of normal distribution of a random variable plays an important role in various fields of science. It is also one of the most common assumptions made in the development and use of the common statistical techniques such as t-test or F-test. It is necessary to verify the assumption of normality when solving practical tasks. Currently, the most popular omnibus test of normality for a general use is the Shapiro-Wilk test. The Jarque-Bera test is the most widely adopted omnibus test of normality in econometrics, finance and related fields. Finally, the Lilliefors test is a representative test based on the comparison of theoretical and empirical distribution function. As outliers in the data sets in the field of economics and finance are frequently present, the Jarque-Bera test is not sufficiently robust, since it is based on the classical characteristics of skewness and kurtosis and has a zero breakdown point. Consequently, the aim of this paper is to derive robust tests of normality that belong to the RT class tests using pseudomedian in the construction of test statistics, and to highlight the benefits of its use in testing normality in the datasets where outliers are present.

Keywords: Location functional, mean, median, pseudomedian, robustness, RT class tests, testing for normality, trimmed mean.

JEL classification: C12, C15, C63
AMS classification: 11K45, 62F03, 65C05, 68U20

1 Introduction

Classic parametric statistical significance tests, such as t-test, analysis of variance and least squares regression, are widely used by researchers in many disciplines, e.g. economics, finance, econometrics, among others. The normality of random variable constitutes one of the most important assumptions in applying these statistical procedures. In other words, the assumption of normality of random variable must be satisfied to produce accurate results of the classic parametric tests. However, the assumption of normality is rarely met when analyzing real financial or economic data sets. The use of classic parametric methods with violated normality assumption may lead to substantive errors in the interpretation of data, because p-values of these tests or confidence intervals for parameters are not credible.

Thus, the testing for normality constitutes one of the most important issues in verifying the correct use of the classic parametric statistical tests. Recently, there are many tests for normality, e.g. the Shapiro-Wilk test (see [12]) as the most popular omnibus test of normality for a general use, the Jarque-Bera test ([7]) as the most widely adopted omnibus test of normality in econometrics, finance and related fields, and finally, the Lilliefors (Kolmogorov-Smirnov) test ([8]) as the most famous test based on the comparison of theoretical and empirical distribution function. Recently, several robust tests for normality have been introduced – see for example [4], [5], [6], [11], [13] and [14].

As it is generally known, the economic and financial data are typical of the occurrence of outliers. As outliers in these data sets are frequently present, the classical Jarque-Bera test is not sufficiently robust, since it is based on the classical characteristics of skewness and kurtosis and has a zero breakdown point[3] (see [3]). Therefore, we need to robustified the classical Jarque-Bera test. Thus, the aim of this paper is to introduce some modified Jarque-Bera tests based on robust location estimators (primarily on pseudo-median) and compare power and robustness of these tests with classical Jarque-Bera test and with other commonly used tests such as Shapiro-Wilk and Lilliefors tests. For the purpose of robustification of the classical Jarque-Bera test, the general RT class tests of normality and modified Jarque-Bera tests using robust location estimators such as median, trimmed mean and pseudo-median will be introduced in next sections.

[1]Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, lubos.strelec@mendelu.cz

[2]Department of Economics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, ladislava.grochova@mendelu.cz

[3]The breakdown point is the proportion of incorrect observations (e.g. arbitrarily large or small extreme values) that must be introduced into sample to cause the estimator to yield an incorrect (e.g. arbitrarily large) result.

2 General RT class

The general RT class is based on robustification of the classical Jarque-Bera test introduced by [7]. The general RT class test statistic is defined by [13] for purpose of robust testing for normality against Pareto tails and has the following general form:

$$RT = \frac{k_1(n)}{C_1}\left(\frac{M_{j_1}^{\alpha_1}(r_1, T_{(i_1)}(s_1))}{M_{j_2}^{\alpha_2}(r_2, T_{(i_2)}(s_2))} - K_1\right)^2 + \frac{k_2(n)}{C_2}\left(\frac{M_{j_3}^{\alpha_3}(r_3, T_{(i_3)}(s_3))}{M_{j_4}^{\alpha_4}(r_4, T_{(i_4)}(s_4))} - K_2\right)^2, \tag{1}$$

where M_j are jth theoretical central moment estimators of the random variable defined as $M_j(r, T(F_n, s)) = \frac{1}{n-2r}\sum_{m=r+1}^{n-r}\varphi_j(X_{(m)} - T(F_n, s))$ for $j \in \{0, 1, 2, 3, 4\}$, where φ_j is a tractable and continuous function, where $\varphi_0(x) = \sqrt{\pi/2}|x|$ and $\varphi_j(x) = x^j$ for $j \in \{1, 2, 3, 4\}$, $X_{(m)}$ is the order statistic, $T(F_n, s)$ is a location functional applied to the sample $X_1, X_2, \ldots, X_n$, r and s are the trimming constants for moments and location, respectively, K_1 and K_2 are small-sample variants of mean corrections, C_1 and C_2 are asymptotic constants, α_1, α_2, α_3 and α_4 are exponents, and finally, $k_1(n)$ and $k_2(n)$ are functions of sample size n. Notice that the general RT class test statistic has asymptotically χ^2-distribution – for more details of theoretical results see [13] and [14].

3 Classical and modified Jarque-Bera tests

As it was noted in [11], the classical Jarque-Bera test statistic is a special case of the general RT class test statistic without trimming for the following parameters: $K_1 = 0$, $K_2 = 3$, $k_1(n) = n$, $k_2(n) = n$, $C_1 = 6$, $C_2 = 24$, $\alpha_1 = 1$, $\alpha_2 = 3/2$, $\alpha_3 = 1$, $\alpha_4 = 2$, $T_{(i_1)} = T_{(i_2)} = T_{(i_3)} = T_{(i_4)} = T_{(0)}$, $j_1 = 3$, $j_2 = 2$, $j_3 = 4$, $j_4 = 2$. Thus, the classical Jarque-Bera test statistic has the form:

$$JB = \frac{n}{6}\left(\frac{M_3\left(0, T_{(0)}(F_n, 0)\right)}{M_2^{3/2}\left(0, T_{(0)}(F_n, 0)\right)}\right)^2 + \frac{n}{24}\left(\frac{M_4\left(0, T_{(0)}(F_n, 0)\right)}{M_2^2\left(0, T_{(0)}(F_n, 0)\right)} - 3\right)^2, \tag{2}$$

which also could be rewritten to the classical form (see [7]):

$$JB = \frac{n}{6}\left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}\right)^2 + \frac{n}{24}\left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3\right)^2. \tag{3}$$

As it is evident from eq. (2), the classical Jarque-Bera test is based on mean ($T_{(0)} = \frac{1}{n}\sum_{i=1}^{n} X_i$ in the general RT class test statistic in eq. (1)) and thus it has asymptotic zero breakdown point (see [3]). However, due to typical presence of outliers in economic and financial data we need the robust test, i.e. test with positive breakdown point. For this purpose, we use the robust Bickel-Lehmann construction of location. This approach has been introduced by P. E. Bickel and E. L. Lehmann in a series of papers (e.g. [2]) and as was shown in [13] and [14] it looks to be playing a crucial role also by robust testing for normality. In this paper we compare the following robust Bickel-Lehmann location estimators with a primary focus on the pseudo-median and its usefulness in normality testing:

- median: $T_{(1)} = F_n^{-1}(1/2)$;
- trimmed mean: $T_{(2)}(s) = \frac{1}{n-2s}\sum_{i=s+1}^{n-s} X_{(i)}$, where $X_{(i)}$ is the i-th order statistic of the sample and $s = 0.05n$ is the trimming constant for location, i.e. we assume, that the smallest 5% of the observations are removed along with the largest 5% of the observations;
- pseudo-median: $T_{(3)} = \text{median}_{i \le j}(X_i + X_j)/2$, i.e. the median of the set $\{(X_1+X_1)/2, (X_1+X_2)/2, (X_1+X_3)/2, \ldots, (X_1+X_n)/2, (X_2+X_2)/2, (X_2+X_3)/2, \ldots, (X_2+X_n)/2, \ldots, (X_{n-1}+X_n)/2, (X_n+X_n)/2\}$.

The mentioned location estimators have different asymptotic breakdown point (see [9]). As it is generally known, the higher breakdown point of an estimator, the more robust it is. The asymptotic breakdown point of mean is zero, on the other hand the asymptotic breakdown point of median is one-half, which is the maximum of breakdown point. The asymptotic breakdown point of trimmed mean is equal to relative expression of trimmed observations (i.e. in our case of 5% trimming the asymptotic breakdown point is 0.05), and finally the asymptotic breakdown point of pseudo-median is $1 - \frac{1}{\sqrt{2}} = 0.293$. Therefore, the pseudo-median represents the compromise between the least robust location estimator (mean) and the most robust location estimator (median) and has great potential in normality testing, because the trade off between power and robustness of tests is very discussed issue in recent literature.

For this reason, we will replace the classical location estimator (mean) in Jarque-Bera test statistic with the robust location estimators (median, trimmed mean and pseudo-median) and compare power and robustness of these modified Jarque-Bera tests. However, for the construction of the modified Jarque-Bera test we need calculate the appropriate constants C_1, C_2, K_1 and K_2 used in general RT class test statistic. As was noted in [14] choosing of appropriate constants is the hardest aspect of the variants of RT class tests, because to obtain the constants C_1 and C_2 we need to find the expressions for $E(M^k_{n_1,n_2})$ for a finite sample size. Such calculations are very tedious and therefore we obtained these constants from Monte Carlo simulations. The results of simulations of constants C_1, C_2, K_1 and K_2 for the analyzed modified Jarque-Bera tests are part of formulas of their test statistics – see eq. (4), (5) and (6).

Now we can introduce the modified Jarque-Bera test statistics based on median, trimmed mean and pseudo-median, respectively:

1. **"Median-Jarque-Bera" test** (MJB), which is a special case of the general RT class test statistic without trimming for the following parameters: $K_1 = 0$, $K_2 = 3$, $k_1(n) = n$, $k_2(n) = n$, $C_1 = 18$, $C_2 = 24$, $\alpha_1 = 1$, $\alpha_2 = 3/2$, $\alpha_3 = 1$, $\alpha_4 = 2$, $T_{(i_1)} = T_{(i_2)} = T_{(i_3)} = T_{(i_4)} = T_{(1)}$, $j_1 = 3$, $j_2 = 2$, $j_3 = 4$, $j_4 = 2$. Thus, the MJB test statistic has the form:

$$MJB = \frac{n}{18}\left(\frac{M_3\left(0, T_{(1)}(F_n, 0)\right)}{M_2^{3/2}\left(0, T_{(1)}(F_n, 0)\right)}\right)^2 + \frac{n}{24}\left(\frac{M_4\left(0, T_{(1)}(F_n, 0)\right)}{M_2^{2}\left(0, T_{(1)}(F_n, 0)\right)} - 3\right)^2. \tag{4}$$

2. **"Trimmed-mean-Jarque-Bera" test (**$TMJB$**)**, which is a special case of the general RT class test statistic with trimming for the following parameters: $K_1 = 0$, $K_2 = 3$, $k_1(n) = n$, $k_2(n) = n$, $C_1 = 8.5$, $C_2 = 24$, $\alpha_1 = 1$, $\alpha_2 = 3/2$, $\alpha_3 = 1$, $\alpha_4 = 2$, $T_{(i_1)} = T_{(i_2)} = T_{(i_3)} = T_{(i_4)} = T_{(2)}$, $j_1 = 3$, $j_2 = 2$, $j_3 = 4$, $j_4 = 2$. Thus, the $TMJB$ test statistic has the form:

$$TMJB = \frac{n}{8.5}\left(\frac{M_3\left(0, T_{(2)}(F_n, 0)\right)}{M_2^{3/2}\left(0, T_{(2)}(F_n, 0)\right)}\right)^2 + \frac{n}{24}\left(\frac{M_4\left(0, T_{(2)}(F_n, 0)\right)}{M_2^{2}\left(0, T_{(2)}(F_n, 0)\right)} - 3\right)^2. \tag{5}$$

3. **"Pseudo-median-Jarque-Bera" test** ($PMJB$), which is a special case of the general RT class test statistic without trimming for the following parameters: $K_1 = 0$, $K_2 = 3$, $k_1(n) = n$, $k_2(n) = n$, $C_1 = 9$, $C_2 = 24$, $\alpha_1 = 1$, $\alpha_2 = 3/2$, $\alpha_3 = 1$, $\alpha_4 = 2$, $T_{(i_1)} = T_{(i_2)} = T_{(i_3)} = T_{(i_4)} = T_{(3)}$, $j_1 = 3$, $j_2 = 2$, $j_3 = 4$, $j_4 = 2$. Thus, the $PMJB$ test statistic has the form:

$$PMJB = \frac{n}{9}\left(\frac{M_3\left(0, T_{(3)}(F_n, 0)\right)}{M_2^{3/2}\left(0, T_{(3)}(F_n, 0)\right)}\right)^2 + \frac{n}{24}\left(\frac{M_4\left(0, T_{(3)}(F_n, 0)\right)}{M_2^{2}\left(0, T_{(3)}(F_n, 0)\right)} - 3\right)^2. \tag{6}$$

4 Power comparison – simulation study

4.1 Simulation setup

In this paper, we perform the Monte Carlo simulations study using R [10] to evaluate the power and robustness of the classical and modified Jarque-Bera tests in testing if a random sample of n independent observations come from a population with a normal distribution. For the simulation study the significance levels $\alpha \in \{0.01, 0.05, 0.10\}$, sample sizes $n \in \{20, 100, 400, 1000\}$ and broad scale of symmetric and asymmetric alternative distributions were considered. However, due to space constrains the results for $\alpha = 0.05$, $n \in \{20, 100\}$ and several alternative distributions will be presented in this paper. The number of replications of Monte Carlo simulations is 100,000.

For the purpose of comparison we simulate the power of the classical Jarque-Bera test (JB), Shapiro-Wilk test (SW), Lilliefors test (LT) as well as three modified Jarque-Bera tests described above (MJB, $TMJB$ and $PMJB$) and robust Jarque-Bera test (RJB) introduced by [5].

As alternative distributions we consider the following distributions: standard Cauchy distribution, Student t(5) distribution, exponential distribution, standard log-normal distribution, two bimodal distributions and four versions of p-location-outliers models introduced by [1]. As bimodal distribution we consider mixture of two normal distributions with distribution function $F = (1-p)N(\mu_1, \sigma_1^2) + pN(\mu_2, \sigma_2^2)$ for $p = 0.50$, $\mu_1 = 0$, $\mu_2 = 3$, $\sigma_1^2 = \sigma_2^2 = 1$, i.e. $F = 0.50N(0,1) + 0.50N(3,1)$. Finally, for the p-location-outliers models we consider $y_1, .., y_{n-p}$ to be *iid* from $N(0,1)$ and $y_{n-p+1}, .., y_n$ to be *iid* from $N(\lambda, 1)$. For our simulation study we consider the following parameters setup: $p \in \{1, 5\}$ and $\lambda \in \{3, 5\}$.

4.2 Results of simulation study

The results of our Monte Carlo simulations are summarized in Tab. 1 and Tab. 2. Based on presented results we can conclude that the most powerful test against symmetric heavy-tailed alternatives such as Cauchy and t(5) is the robust Jarque-Bera (RJB) test. If we suppose asymmetric light- and heavy-tailed alternatives such as exponential and log-normal distributions as well as short-tailed symmetric alternatives such as uniform and bimodal distributions we can conclude that the most powerful test is the Shapiro-Wilk (SW) test. If we compare the three modified Jarque-Bera tests described above we can conclude that pseudo-median-Jarque-Bera ($PMJB$) test is suitable for testing of normality against heavy-tailed symmetric alternatives and median-Jarque-Bera (MJB) test is suitable for testing of normality against heavy- and light-tailed asymmetric alternatives. Furthermore, we can state that all tests based on the Jarque-Bera test statistic are biased against uniform and bimodal alternatives in case of small sample size $n = 20$, RJB test even in case of large sample size $n = 100$.

	Cauchy	t(5)	exp(1)	lnorm	unif
LT	0.845	0.134	0.587	0.794	0.096
SW	0.863	0.188	0.826	0.932	0.193
JB	0.861	0.234	0.627	0.829	0.003
RJB	0.902	0.242	0.557	0.785	0.002
MJB	0.857	0.221	0.670	0.858	0.005
$TMJB$	0.852	0.232	0.616	0.819	0.002
$PMJB$	0.858	0.229	0.644	0.839	0.005
			p-outlier models		
	bimodal	$p = 1$		$p = 5$	
	$\mu = 3$	$\lambda = 3$	$\lambda = 5$	$\lambda = 3$	$\lambda = 5$
LT	0.087	0.122	0.419	0.243	0.824
SW	0.098	0.231	0.772	0.246	0.882
JB	0.001	0.319	0.874	0.060	0.129
RJB	0.000	0.306	0.855	0.059	0.149
MJB	0.001	0.297	0.848	0.091	0.284
$TMJB$	0.001	0.322	0.876	0.050	0.046
$PMJB$	0.001	0.307	0.862	0.091	0.416

Table 1 Power of the selected classical and robust normality tests for $\alpha = 0.05$ and small sample size $n = 20$

If we focus on outliers models, we can conclude that the most commonly used tests of normality, such as the SW and JB tests, are too strict in rejecting normality in the case of a small number of outliers, even when the sample size is large enough. To illustrate this general framework, notice that one of the most used SW test rejects the hypothesis of normality in approximately 80% cases if we suppose the p-location-outlier model for $p = 1$, $\lambda = 5$ and sample size $n = 100$. Similarly, all of analyzed versions of the JB tests are too strict in rejecting normality for the same p-location-outlier model and sample size – all of these tests reject the hypothesis of normality in more than 85% of cases. From this point of view the most robust test in our study is LT test which rejects the hypothesis of normality in approximately 18%.

If we suppose a larger number of outliers in small sample size – e.g. the p-location-outlier model for $p = 5$, $\lambda = 5$ and sample size $n = 20$ – we need the test with high power, because the hypothesis of normality is not sustainable. Based on our Monte Carlo simulations we can recommend the SW and LT tests – these test reject the hypothesis of normality in approximately 88% and 82% cases, respectively. In contrast, all of analyzed versions of the JB tests have small power against the mentioned p-location-outlier model, e.g. classical JB test rejects the hypothesis of normality only in approximately 13% cases. However, if we compare the power of the classical JB test with power of the pseudo-median-Jarque-Bera ($PMJB$) test we can observe a significant increase in power in the case of $PMJB$ test – this test rejects the hypothesis of normality in approximately 42% cases, which is approximately double value compared to other versions of JB tests.

	Cauchy	t(5)	exp(1)	lnorm	unif
LT	1.000	0.327	1.000	1.000	0.583
SW	1.000	0.566	1.000	1.000	0.997
JB	1.000	0.644	1.000	1.000	0.738
RJB	1.000	0.671	1.000	1.000	0.005
MJB	1.000	0.638	1.000	1.000	0.758
$TMJB$	1.000	0.643	1.000	1.000	0.679
$PMJB$	1.000	0.640	1.000	1.000	0.579
			p-outlier models		
	bimodal	$p = 1$		$p = 5$	
	$\mu = 3$	$\lambda = 3$	$\lambda = 5$	$\lambda = 3$	$\lambda = 5$
LT	0.615	0.072	0.180	0.336	0.985
SW	0.734	0.194	0.804	0.681	1.000
JB	0.209	0.267	0.881	0.740	1.000
RJB	0.001	0.248	0.858	0.742	1.000
MJB	0.259	0.256	0.867	0.702	1.000
$TMJB$	0.171	0.267	0.880	0.747	1.000
$PMJB$	0.121	0.262	0.875	0.742	1.000

Table 2 Power of the selected classical and robust normality tests for $\alpha = 0.05$ and large sample size $n = 100$

5 Discussion and summary

In this paper we introduce three robust modifications of the classical Jarque-Bera test based on three robust location estimators – median, trimmed mean and pseudo-median. The main focus was on pseudo-median and its usefulness in normality testing, because the pseudo-median has good properties, especially from the point of view of breakdown point – as was presented in this paper, the asymptotic breakdown point of pseudo-median is $1 - \frac{1}{\sqrt{2}} = 0.293$, while the asymptotic breakdown point of mean is zero and the asymptotic breakdown point of median is one-half. Therefore, the pseudo-median represents the good compromise between the least robust location estimator (mean) and the most robust location estimator (median).

Based on our simulation study we can conclude that the pseudo-median-Jarque-Bera ($PMJB$) test is suitable for testing of normality against heavy-tailed symmetric alternatives and also significantly increases the power of the classical Jarque-Bera test in the case of presence of a larger number of outliers in small sample size data set. Finally, we can state that the pseudo-median has great potential in normality testing, because the trade off between power and robustness of tests is very discussed issue in current literature focusing on normality testing.

The suggested modifications of the classical Jarque-Bera test could be useful for use in the applied regression analysis, because testing for normality of the error terms constitutes one of the most important steps of regression model verification and validation. Specifically, these robust tests are suitable for use in case of occurrence of outliers in the data sets – for more see e.g. [15] and [16].

Acknowledgements

Supported by the grant No. GA16-07089S of the Czech Science Foundation.

References

[1] Balakrishnan, N.: Permanents, Order Statistics, Outliers, and Robustness. *Revista Matemática Complutense*, **20**, 1 (2007), 7–107.

[2] Bickel, P. J., and Lehmann, E. L.:. Descriptive Statistics for Nonparametric Models II. Location. *Annals of Statistics*, **3**, 5 (1975), 1045–1069.

[3] Brys, G., Hubert, M., and Struyf, A.: A robustification of the Jarque-Bera test of normality. In: *COMPSTAT 2004, Proceedings in Computational Statistics*, Springer, Physica Verlag, 2004, 753–760.

[4] Brys, G., Hubert, M., and Struyf, A.: Goodness-of-fit tests based on a robust measure of skewness. *Computational Statistics*, **23**, 3 (2008), 429–442.

[5] Gel, Y. R., and Gastwirth, J. L.: A robust modification of the Jarque-Bera test of normality. *Economics*

Letters, **99**, 1 (2008), 30–32.

[6] Gel, Y. R., Miao, W., and Gastwirth, J. L.: Robust directed tests of normality against heavy-tailed alternatives. *Computational Statistics & Data Analysis*, **51**, 5 (2007), 2734–2746.

[7] Jarque, C. M., and Bera, A. K.: A test for normality of observation and regression residuals. *International Statistical Review*, **55**, 2 (1987), 163–172.

[8] Lilliefors, H.: On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, **62**, 318 (1967), 399–402.

[9] Maronna, R. A., Martin, R. D., and Yohai, V. J.: *Robust statistics: Theory and methods*. John Wiley & Sons, Ltd., Chichester, 2006.

[10] R Core Team: *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016, http://www.R-project.org/.

[11] Richter, W.-D., Střelec, L., Ahmadinezhad, H., and Stehlík, M.: Geometric aspects of robust testing for normality and sphericity. *Stochastic Analysis and Applications* **35**, 3 (2017), 511–532.

[12] Shapiro S. S., Wilk M. B.: An analysis of variance test for normality. *Biometrika*, **52**, 3–4 (1965), 591–611.

[13] Stehlík, M., Fabián, Z., and Střelec, L.: Small sample robust testing for normality against Pareto tails. *Communications in Statistics – Simulation and Computation* **41**, 7 (2012), 1167–1194.

[14] Stehlík, M., Střelec, L., and Thulin, M.: On robust testing for normality in chemometrics. *Chemometrics and Intelligent Laboratory Systems* **130**. (2014), 98–108.

[15] Střelec, L., Grochová, L, and Kolman, P.: On power comparison of normality tests of error terms in regression models. In: *11th International Conference of Numerical Analysis and Applied Mathematics 2013: ICNAAM 2013*, edited by T. E. Simos, et al., AIP Conference Proceedings 1558, American Institute of Physics, New York, 2013, 1847–1850. ISBN 978-0-7354-1184-5.

[16] Střelec, L.: Notes on power of normality tests of error terms in regression models. In: *12th International Conference of Numerical Analysis and Applied Mathematics 2014: ICNAAM 2014*, edited by T. E. Simos, et al., AIP Conference Proceedings 1648, American Institute of Physics, New York, 2015, 720013(1)–720013(4). ISBN 978-0-7354-1287-3.

The dynamic aspects of income in terms of consumption function in EU countries

Kvetoslava Surmanová[1], Zlatica Ivaničová[2], Marian Reiff[3]

Abstract This paper deals with modeling of aggregate household consumption at the macro level. We focused on the examination of the impact of the dynamic aspect of wages for consumption in selected countries of the European Union. The main motivation for this analysis was to examine the differences between countries in the labor market. Dynamic models allow you to analyze how consumption in each country responds to wage delays. Partial attention is focused on short-term and long-term multipliers consumption function in different countries. Based on empirical research it is possible in the analyzed countries in terms of the dynamic impact of wages on consumption to identify a group of countries in which the amount of wages creates sufficient space for saving a group where the valuation work undervalued and the formation of savings occurs at a low level. The analysis used data from Eurostat from 1996 to 2015.

Keywords: Consumption function, wages, dynamic model, ADL, households.

JEL Classification: E21
AMS Classification: 62P20

1 Introduction

Household consumption is the main component of aggregate demand and, therefore is important in the economic analysis and in particular in macroeconomics. Among the most common and best-known approaches to the analysis of consumption function can be included Keynesian theory. The main well-known basic characteristics of Keynes´ analysis are that the marginal propensity to consume (MPC) falls with income, as does the average propensity to consume (APC).

In response to the Keynes´ theory Milton Friedman proposed in 1956 his permanent income hypothesis. The main idea of the permanent income hypothesis of Friedman arises from the fact that households are faithful to their consumption patterns, regardless of short-term income and thus consumption is much more stable than expected. According to Friedman, population realizes its consumption with respect to the amount of permanent income[4].

At the same time that Friedman was developing theory of lifecycle model. Authors were also Modigliani and Brumberg in 1955. According to lifecycle theory residences maximize their lifetime utility subject to their lifetime budget constraint. It is most often formulated at the macro level, but we can also see it applied in microeconomics.

Both theories (permanent-income hypothesis and life-cycle hypothesis) were developed at least partly in response to the discrepancy between Keynes's conjecture and the empirical evidence.

The theory of consumer function has since been associated with empirical research in this area. The individual theoretical components of consumer design were linked to the development of econometrics. Problematic autocorrelation of the random component in consumer function can be solved by incorporating the dynamics into the model. Such a model can be constructed as an ADL model (Autoregressive Distributed Lag model). This type of model is also in line with the theoretical starting points of Friedman [5].

Davidson, Hendry, Srba and Yeo [4] published in 1978 theory based on Friedman´s hypothesis. It was the first econometric model of consumption function incorporating an error correction mechanism (ECM).

[1] University of Economics, Dolnozemská 1, Bratislava, Slovakia, e-mail: kvetoslava.surmanova@euba.sk
[2] University of Economics, Dolnozemská 1, Bratislava, Slovakia, e-mail: zlatica.ivanicova@euba.sk
[3] University of Economics, Dolnozemská 1, Bratislava, Slovakia, e-mail: marian.reiff@euba.sk
[4] Permanent mean to be oriented long-term increase in income.

Many studies have used these models or a combination of the variables suggested by these approaches, in order to explain consumption (see Byrne and Davis [3], Bover [2], Senaj [6], Arlt et al. [1]). Consumption is also part of simultaneous and structural models (see Szomolányi [7]).

The aim of this paper is to verify the relationship between the consumption of households and wages in selected EU countries for the period 1995 - 2015. In our analysis, we use the ADL model.

The rest of the paper is organized as follows: section 2 deals with general specification of ADL model and long-term multiplier, section 3 provide a description of the data and estimation results and the last section concludes.

2 Concepts of consumption function econometric modeling

As a reaction to Keynes's theory, further advanced theories of consumption and savings have been developed. Dynamics is a fundamental feature of the macroeconomy and all serious policy analysis is based on a dynamic approach. Let us consider the model in the form ADL(*p*, *q*):

$$c_t = \beta_0 + \delta_0 w_t + \delta_1 w_{t-1} + \delta_2 w_{t-2} + \cdots + \delta_q w_{t-q} + \beta_1 c_{t-1} + \beta_2 c_{t-2} + \cdots + \beta_p c_{t-p} + u_t, \quad (1)$$

where c_t is a final consumption of households in current prices (million €) and w_t are wages in current prices (million €). Such a dynamic model is characterized by two features. The first is that it much better explains the dependent variable than the static version. The second is that we can get a dynamic multiplier by its estimate. The short-run multiplier δ_0 expresses the instant response of the dependent variable to the unit change of the independent variable. The long-run multiplier expresses the equilibrium value of the multiplier in the long run and we can express it from the general ADL model by the form:

$$(\delta_0 + \delta_1 + \delta_2 + \cdots \delta_q)/(1 - \beta_1 - \beta_2 - \cdots - \beta_p), \quad (2)$$

under the conditions $-1 < \beta_i < 1$.

3 Data and empirical results

The data used in this study were retrieved from the Eurostat database [8]. The used variables are wages and consumption of households from 1995 to 2015 (defined at current market prices in mil. Euro). Our dataset covers 14 countries from all 28 countries of EU: CY – Cyprus, CZ – Czech Republic, EE – Estonia, IR – Ireland, El – Greece, ES – Spain, LV – Latvia, LT – Lithuania, HU – Hungary, MT – Malta, PL – Poland, PT – Portugal, SL – Slovenia and SK – Slovakia. The time span 1995 – 2015 was chosen because there were no reliable data available for all analyzed countries before 1995 and the year 2015 was the last year of published statistics by Eurostat. The whole analysis was carried out in the software Eviews.

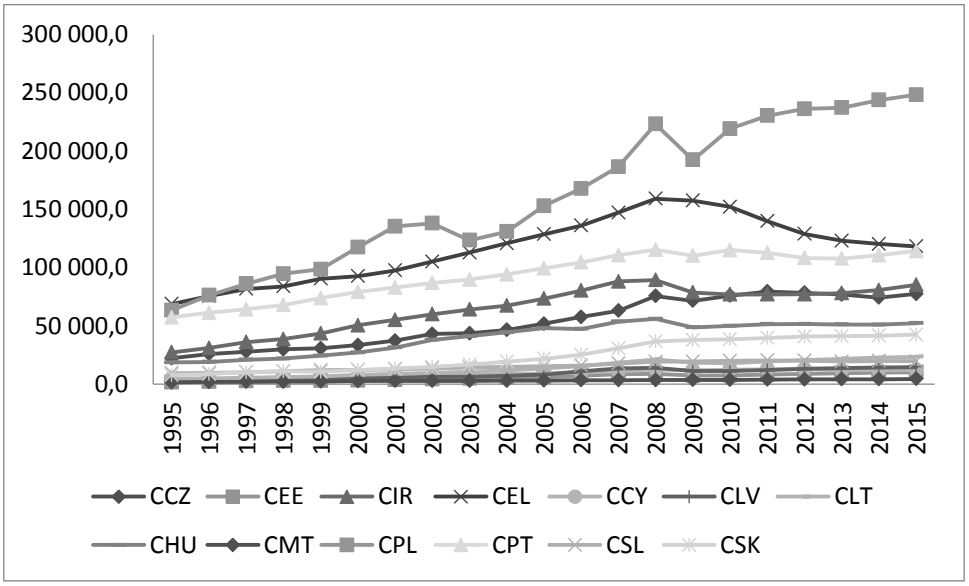


Figure 1 Final household consumption in selected EU countries over time

Consumption households in analyzed countries have growth. The most significant increase over the monitored period was in Lithuania and Estonia (almost 6 times) and in Latvia and Slovakia more than 4 times. This

increased consumption is mainly driven by employment growth and wage growth. In all four countries, the increase in aggregate wages was significantly higher than in the rest of the countries (Latvia - 7.1 times, Lithuania - 6.8 times, Estonia - 5.2 times and Slovakia - 4.1 times). Graphical consumption patterns of households (Figure 1) do not show the consumption of Spain, as its values significantly exceed the level in other analyzed countries. This heterogenity can also be seen in Table 1.

	CZ	***CY***	***EE***	***EL***	***ES***	***HU***	***IR***
Mean	52 229,79	8 977,67	5 746,79	116 054,92	482 747,71	39 788,32	63 638,74
Median	49 161,30	8 789,25	5 669,80	120 485,85	509 877,50	45 895,20	70 415,85
Minimum	22 178,60	4 603,40	1 516,50	68 737,80	281 496,50	18 681,20	27 241,80
Maximum	79 380,40	12 806,60	9 818,20	159 108,20	623 029,00	55 967,00	89 368,10
	LV	***LT***	***MT***	***PL***	***PT***	***SL***	***SK***
Mean	8 565,51	13 224,19	3 063,55	157 610,31	92 576,18	15 371,56	23 916,83
Median	7 719,90	12 688,70	3 073,25	145 414,35	96 801,10	15 101,15	20 606,30
Minimum	2 697,90	3 267,70	1 659,70	63 399,50	57 376,20	9 431,20	7 695,90
Maximum	14 165,80	22 761,90	4 255,00	243 631,70	115 216,20	20 337,90	41 605,20

Table 1 Descriptive statistics of consumption of households

It follows that the concept of consumer service should have a high degree of adequacy. Based on our experiments, we have compiled an autoregressive model of consumption for each selected country. Differences between countries were recorded in an adequate and statistically significant impact of wage delays. The results are shown in the following table.

country	constant	c(t-1)	w(t)	w(t-1)	R^2	propensity to consume	model	long-run multiplier
CY	0,78***	0,48***	0,45***		0,988	0,45	ADL(1, 0)	0,87
CZ	0,49***	0,72***	0,81***	-0,57***	0,999	0,24	ADL(1, 1)	0,86
EE	0,54***	0,59***	0,99***	-0,63***	0,997	0,36	ADL(1, 1)	0,88
EL	1,34**	0,63***	0,56***	-0,28*	0,991	0,28	ADL(1, 1)	0,76
ES	0,41	0,85***	0,69***	-0,57***	0,994	0,12	ADL(1, 1)	0,80
IR	1,16***	0,26	0,98***	-0,34*	0,993	0,64	ADL(1, 1)	0,86
LV	0,35	0,9***	0,61***	-0,55***	0,992	0,06	ADL(1, 1)	0,60
HU	0,61**	0,58**	0,98***	-0,61***	0,996	0,37	ADL(1, 1)	0,88
LT	0,51**	0,66***	0,76***	-0,45**	0,996	0,31	ADL(1, 1)	0,91
MT	0,53***	0,87***	0,73***	-0,66**	0,993	0,07	ADL(1, 1)	0,54
PL	0,71**	0,66***	0,83***	-0,53***	0,994	0,30	ADL(1, 1)	0,88
PT	-0,23	0,87***	0,77***	-0,61***	0,992	0,16	ADL(1, 1)	1,23
SK	-0,03	0,26	0,78***		0,994	0,78	ADL(1, 0)	1,05
SL	0,67***	0,4***	0,54***		0,996	0,54	ADL(1, 0)	0,90

Table 2 Results of estimated ADL consumption function, own calculations
***, **, * Parameter is significant on 1, 5 and 10 % level of significance

Figure 2 shows development propensity to consume and impact of consumption analyzed countries. The high value of estimated parameters of consumption in time $t-1$ was in countries Latvia (0,9), Malta and Portugal (0,87), Estonia (0,85) and Czech republic (0,72). In this countries are high inertia in household behavior. On the other hand, the lowest values were measured in Slovakia and Ireland (both 0,26) reflect low inertia. It means that consumption is determined largely by the income of the population. Using the ADL(1, 0) or ADL(1, 1) models, countries can be divided into two groups, on in which the wages of the population are consumed in the relevant period. This group includes Cyprus, Slovakia and Slovenia. In these countries the results point to the fact that,

households have too low income to makes savings. In the remaining 11 countries is consumption determined by wages for the past period, or we could interpret it as the impact of savings.

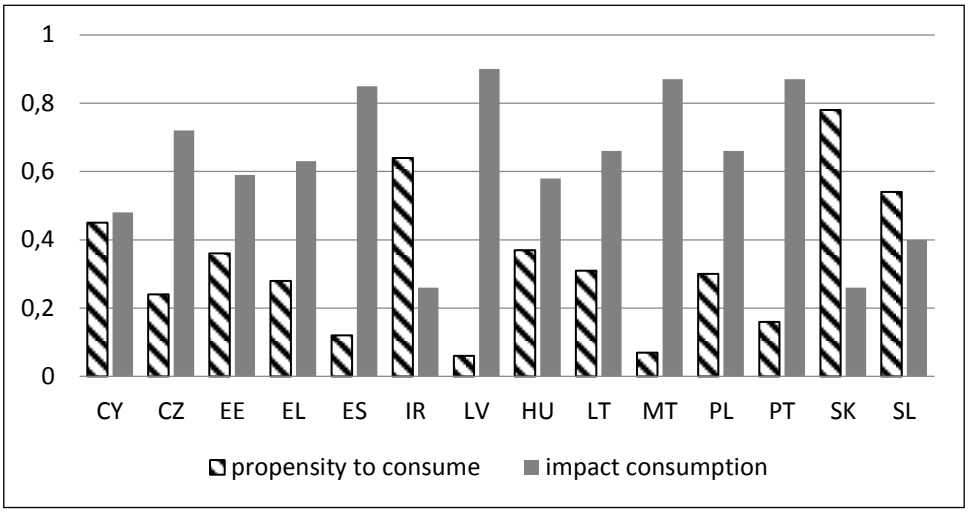


Figure 2 Multiplier, Propensity to consume and impact consumption in selected EU countries

The values of the finally indicators are not very different, there are only small exceptions. According to the value of the long-run multiplier, the impact of wage growth in the long run has the most significant impact on the consumption of the population in two countries - Slovakia and Portugal. Increase in wages by 1% will cause an increase in consumption by more than 1%. On the contrary, the lowest impact of the results is shown by the households Latvia and Malta.

4 Conclusion

The aim of this paper was to specify and estimate a dynamic consumption function. We focus on detection impact of consumption time ago and on long-run multiplier of wages.

Frist we estimated two version of ADL models, ADL(1, 0) and ADL(1, 1) according to significant lag of variable. Next we calculated propensity to consume and multiplier. We used graphical methods to compare results.

From the results (Figure 2) we can conclude that in Spain, Latvia, Malta, Portugal, Czech Republic and Greece the consumption of households is determined with inertia in consumption and the impact growth of wages is the lowest form the analyzed countries. The difference between consumption inertia and propensity is not so significant in Cyprus, Estonia, Hungary, Lithuania and Poland. The opposite effect was demonstrated in three countries: Slovakia, Ireland and Slovenia.

A high propensity to consume in Slovakia, Slovenia and Ireland means that wages are low and cover current consumption.

From the long-term view, the impact of wages on the costumer behavior is balanced in analyzed countries. Vice versa in Slovakia and Portugal growth of wages is much stronger as in other analyses countries.

Differences in the countries of European Union can be seen from different point of view. For example: population expectation on consumption, expectations of future income, or consumer habits or savings.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic - VEGA, grant no. 1/0245/15 and 1/0351/17.

References

[1] Arlt, J., Čutková, J.and Radkovský, Š. 2001. *Analýza spotřební funkce v podmínkach ČR*. WP **34**, Praha 2001.

[2] Bover, O.: *Wealth effects on consumption: Microeconometric estimates from the Spanish survey of household finances*. Bank of Spain working paper, Documentos de Trabajo, No. **0522**, 2005.

[3] Byrne, J. P. and Davis, P. E.: Disaggregate wealth and aggregate consumption: an investigation of empirical relationships for the G7. *Oxford Bulletin of Economics and Statistics* **65** (2), 2003, 197 – 220.

[4] Davidson, J. E. H., Hendry, D. F., Srba, F.and Yeo, S.: Econometric Modelling of the Aggregate Time-Series Relationship Between consumers' Expenditure and Income in the United Kingdom. *The Economic Journal* **88** (1978), 661 – 692.
[5] Friedman, M. A.: *A Theory of the Consumption Function. Princeton*, Princeton University Press, 1957.
[6] Senaj, M.: *Odhad spotrebnej funkcie pre Slovensko a prognóza spotreby*. WP 1/2007, 2007. Retrieved from www.nbs.sk/PUBLIK.
[7] Szomolányi, K., Lukáčik, M., Lukáčiková, A.: Impact of terms-of-trade on Slovakia, the Czech Republic, and Croatia in the short run. In: *Our economy: Journal of Contemporary Issues in Economics and Business. - Maribor*: Faculty of Economics and Business 63 (2017), 3-13.
[8] Eurostat. Retrieved from http://ec.europa.eu/eurostat/data/database.

Using of Markov chains with varying state space for predicting short-term of the share price movements

Milan Svoboda[1], Mikuláš Gangur[2]

Abstract. This paper deals with stochastic modelling and short time prediction of a share price. It follows the works that use Markov chains analysis (MCA) with unvarying state space for predicting the share price movement. According to this analysis business strategies for the purchase and subsequent sale of shares were created. These strategies outperformed the market represented by the passive strategy Buy and Hold. This study uses MCA with varying state space. The state space is defined parametrically as a multiple of moving standard deviation. Three models of state space are calculated. The state space is defined by a moving standard deviation of lengths 10, 20, 30. Nine trading strategies are calculated for each of the models. For each of these trading strategies the achieved yield and the number of transactions are calculated. The study was performed on historical daily prices (open and close) of the CEZ shares in the ten years period from early 2006 to the end of 2015. The results are compared with the strategies which use MCA with unvarying state space and the passive strategy Buy and Hold.

Keywords: Markov chains analysis, technical analysis indicators, share price predication.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Technical analysis (TA) is one of the possibilities used for the prediction of the price development. TA is understood as a large set of methods which can predict future prices from the past ones. The principles and methods of TA are widely described, for example, in [6]. The thesis that share quotations move within trends that have certain inertia is one of the bases of TA. The change of trend is given by the change of the relation between seller and buyer. It is necessary to identify these changes of trend immediately by studying historical prices and volumes of trades. Technical analysts distinguish three kinds of trends. These are as follows: a primary trend lasting from one year up to a few years, a secondary trend lasting several months, and a tertiary trend lasting days up to weeks.

Today, technical analysts concentrate on prognostication of short-term price movements in which the important thing is the estimate of price changes, not price level. TA is based on some scientific theories but mainly it is grounded in a number of empiric findings. On the basis of these empiric findings individual methods are created whose number is practically unlimited. These methods can be divided into two large groups:

- graphic methods, based on regularly repeated formations in graphs;
- methods based on technical indicators whose role is to detect the change of trend. When the trend changes, a buying signal (a growing trend begins) or a selling signal (a decreasing trend begins) is generated.

The aim of this study is to verify the possibility of creating a technical indicator using the Markov chains analysis (MCA) with varying state space. The MCA is used very little for a stochastic description of the behaviour of share markets. Infrequent application of MCA is probably also caused by studies [2] and [8]. What can be labelled as the common thread of these studies is the fact that they define the state space very simply. The state space is defined on the basis of the size of daily changes and the state space defined in such a way does not allow for suitable applications. Contrary to the above, the study [6] defines the state space on the basis of the size of daily cumulative changes of the share price. The study shows that in a state space that is defined in such a way there are states in which there is a high chance of a minority trend change. This approach is also the basis of this study. In study [6] individual states are defined as multiples of the standard deviation of daily changes calculated for the entire monitored period. In this study we would like to modify this principle and to define the state space as multiples of the moving standard deviation.

[1] University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic, svobodm@kem.zcu.cz
[2] University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic, gangur@kem.zcu.cz

For the sake of completeness let us state that, for a long time, technical analyses had been ignored by academics. The reason for that is the Efficient Market Theory (EMT) which was formulated by E. Fama [3] on the basis of his studies. According to this theory all the estimated information is already encompassed in the share price. Unexpected information is considered to be the cause of the change of the quotation. Any market that quickly (immediately) absorbs all this unexpected information is understood as effective. Unexpected information is hard to expect, it arrives randomly, and therefore even the change of a share price is random. If this theory was considered valid, TA would be inapplicable and in the long-term all investors would achieve comparable results. With mass introduction of information technology and with application of large volume of data stored in databases, more and more empiric studies occur, for example [7] and [1], and they show that TA methods can overcome the market. The market is usually represented by a passive theory Buy and Hold (B&H), i.e. purchasing a share and holding it passively for all the time.

2 Methodology

2.1 Markov chains

MC theory is described for example in [4]. MC is a random process with a discrete set of states, discrete time and of that kind that the probability $p_i(n)$, that at the moment n the process will be in state i, is stochastically dependent only on the state at the previous time moment, i.e. on the state at the moment n-1. Particular realizations x_i are elements of a countable set $S = \{s_i\}$, $i = 1, 2,...., N$ which is called a state space. Behaviour of the described process is determined by:

- Vector of unconditional probabilities $p(n)^T=[p_1(n), p_2(n), ..., p_N(n)]$, where T means transposition and $p_i(n)$ denotes probability that the process is in the moment n in the state i.
- Transition probability matrix **P** whose elements p_{ij} give conditional probability of process transition from the state i to the state j. That could be formally described $p_{ij} = P(X_n = s_j | X_{n-1} = s_i)$, where $i = 1, 2, .. N$ and $j = 1, 2, ... N$ and where p_{ij} can depend on n. In case that p_{ij} does not depend on n at all we speak about homogenous MC, in the opposite case we speak about non-homogenous MC.

2.2 Data and state space

The study has been carried out on the shares of ČEZ company, and the monitored period of time is from the beginning of year 2006 up to the end of year 2015. The data were provided by company Patria Direct. The company paid dividends in the above period. The paid and reinvested dividends are calculated. Cumulative changes of a share price are calculated from the daily closing prices. A cumulative change of price Y_t is interpreted as short basic indexes of daily closing prices where the day of the change of a minority trend, i.e. the transition from a fall into a growth or vice versa is the basic period. The length of the trend duration is determined by the number of consecutive growing or falling closing prices. Y_t can be calculated according to the following relations:

$$\begin{aligned} Y_t &= Y_{t-1}\frac{P_t}{P_{t-1}} \text{ if } \left(P_{t-2} \le P_{t-1} \le P_t\right) \text{ or} \left(P_{t-2} \ge P_{t-1} \ge P_t\right), \\ Y_t &= \frac{P_t}{P_{t-1}} \text{ otherwise,} \end{aligned} \tag{1}$$

where P_t is the closing daily price at time t, P_{t-1} is the closing daily price at time $t-1$ and P_{t-2} is the closing daily price in time $t-2$. We are going to define the state space on the values y_t which are the percentage expression of Y_t. A set with eight states will be used for data classification. The states when the share price decreases will be labelled D_i. State D_1 will be the state with the smallest decrease of price, and, on the contrary, state D_4 will label the state with the highest decrease of price. The states when the price of share is growing will be labelled G_i. State G_1 will be the state with the smallest growth of price, and, on the contrary, state G_4 will label the state with the highest growth of price. A general model of the state space will be defined on the following principle:

D_4: $y_t < -3s_{t,n}$; D_3: $-3s_{t,n} \le y_t < -2s_{t,n}$; D_2: $-2s_{t,n} \le y_t < -1s_{t,n}$; D_1: $-1s_{t,n} \le y_t < 0$;

G_1: $0 \le y_t < 1s_{t,n}$; G_2: $1s_{t,n} \le y_t < 2s_{t,n}$; G_3: $2s_{t,n} \le y_t < 3s_{t,n}$; G_4: $3s_{t,n} \le y_t$,

where $s_{t,n}$ is the moving standard deviation of length n, which can be calculated according to the following formula:

$$s_{t,n} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}\left(x_{t-i} - \bar{x}_{t,n}\right)^2}, \tag{2}$$

where x_{t-i} is the daily change of the share price on day *t-i* and $\bar{x}_{t,n}$ is the moving average lengths *n* on day *t*. The procedure of the assigning of the states is illustrated in Table 1.

t	P_t	daily change	y_t	$s_{t,10}$	state
30-12-2015	444.3	1.76%	8.18%	1.36%	G_4
29-12-2015	436.6	1.42%	6.31%	1.28%	G_4
28-12-2015	430.5	1.53%	4.82%	1.53%	G_4
23-12-2015	424.0	2.91%	3.24%	1.50%	G_3
22-12-2015	412.0	0.32%	0.32%	1.18%	G_1
21-12-2015	410.7	-0.94%	-3.77%	1.71%	D_3
18-12-2015	414.6	-1.24%	-2.86%	1.71%	D_2
17-12-2015	419.8	-1.22%	-1.64%	1.77%	D_1
16-12-2015	425.0	0.00%	-0.42%	1.95%	D_1
15-12-2015	425.0	-0.42%	-0.42%	2.00%	D_1
14-12-2015	426.8	0.21%	0.21%	2.08%	G_1

Table 1 Procedure for assigning states

On the whole, we are going to create 4 models of state space. Three models with varying state space in which the state space will be defined on the basis of a multiple of the moving standard deviations with lengths 10, 20, and 30. The fourth, non-adaptive model with a non-moving standard deviation (s = 1,84%), will be used as a comparative one and in the next part of the text it will be labelled as basic. The models with varying state space will be labelled according to the length of the moving standard deviation s_{10}, s_{20}, s_{30}. For each model we will calculate 9 (3x3) trading strategies. The buying signal will be gradually generated by states D_2, D_3, D_4 and the selling signals gradually by states G_2, G_3, G_4. The states D_1 a G_1 are not used for generating signals as we require a certain minimum decrease and a minimum growth of the share price.

2.3 Trading strategies

Trading strategies are created on the following principle. When a certain level of a fall of a share price is achieved, a buying signal will be generated and when a certain level of a growth of a share price is achieved, a selling signal will be generated. Trading strategies are always implemented according to the following rules. One trade (transaction) is understood as a share buying transaction and the following one as a share selling transaction. If a buying or selling signal is generated one day, the trade is implemented for the opening price from the following day. The whole capital is always invested (it means that it is theoretically possible to buy a part of a share). We do not take into consideration any transaction fees. We count and reinvest the paid out dividends after tax in case that we obtained the shares on the record day. Dividends are included in the moment of share selling. A short selling is not taken into account and two consecutive purchases are not possible. The invested capital value is calculated according to the following relation:

$$C_n = C_0 \prod_{i=1}^{n} \frac{S_i + D_i}{B_i}, \tag{3}$$

where $C_0 = 1.000$ is the initial value of the capital, C_n is the value of the capital after the n^{th} transaction, S_i is the selling price in the i^{th} transaction, D_i are dividends after taxation in case that during the i^{th} it was the decisive day, B_i is the purchasing price in the i^{th} transaction. To compare the successfulness of the individual models we will still determine the average yield of $\bar{C}$, which can be ascertained according to the below formula:

$$\bar{C} = \frac{C_{D_2-G_2} + C_{D_2-G_3} + \ldots + C_{D_4-G_4}}{9}, \tag{4}$$

where $C_{Di\text{-}Gj}$ is the achieved valuation of the trading strategy in which the buying signal generates state D_i and the selling signal generates state G_j.

3 Results and discussion

First, let us take a look at the probabilities of the transition between the individual states. As well as in [6], first we are going to carry out filtration. What we mean by filtration is the release of consecutive, identical states. By means of filtration we will get rid of the states in which the share price stagnates (or, more precisely, it only changes a little within the same trend). These trends are not interesting from the trading point of view. To illustrate this, let us state a part of the chain before filtration: … D_4, G_1, G_1, G_1, G_2, D_1, D_1, D_2, G_2, D_2, D_2, … and after the filtration … D_4, G_1, G_2, D_1, D_2, G_2, D_2, … . We will determine a probability matrix of transition P for the filtrated MC. The probabilities of the transition are stated in Table 2. The found probabilities are shown only up to two decimal places, and therefore the sum of probabilities may not equal 1 precisely. Direct transitions between some states are impossible, and in these impossible transitions the nought is not followed by any decimal places. Columns $\sum G_i$a $\sum D_i$ state a probability of the remaining in a trend or a change of a trend. The number of occurrences of the individual states is given in the last column.

		D_4	D_3	D_2	D_1	G_1	G_2	G_3	G_4	$\sum D_i$	$\sum G_i$	n
D_4	*basic*	0	0	0	0	0.36	0.45	0.12	0.07	**0.00**	**1.00**	58
	s_{10}	0	0	0	0	0.51	0.42	0.07	0.00	**0.00**	**1.00**	101
	s_{20}	0	0	0	0	0.53	0.31	0.15	0.01	**0.00**	**1.00**	96
	s_{30}	0	0	0	0	0.57	0.22	0.17	0.04	**0.00**	**1.00**	90
D_3	*basic*	0.30	0	0	0	0.55	0.12	0.03	0.00	**0.30**	**0.70**	93
	s_{10}	0.39	0	0	0	0.42	0.17	0.01	0.00	**0.39**	**0.61**	157
	s_{20}	0.42	0	0	0	0.39	0.17	0.02	0.00	**0.42**	**0.58**	142
	s_{30}	0.38	0	0	0	0.40	0.19	0.02	0.00	**0.38**	**0.62**	141
D_2	*basic*	0.08	0.25	0	0	0.55	0.10	0.01	0.00	**0.33**	**0.67**	260
	s_{10}	0.11	0.31	0	0	0.37	0.18	0.03	0.00	**0.42**	**0.58**	303
	s_{20}	0.09	0.30	0	0	0.43	0.15	0.03	0.00	**0.38**	**0.62**	298
	s_{30}	0.09	0.26	0	0	0.45	0.15	0.04	0.00	**0.36**	**0.64**	272
D_1	*basic*	0.01	0.03	0.31	0	0.55	0.08	0.01	0.00	**0.35**	**0.65**	483
	s_{10}	0.01	0.08	0.32	0	0.42	0.15	0.02	0.00	**0.41**	**0.59**	405
	s_{20}	0.02	0.05	0.34	0	0.44	0.13	0.02	0.00	**0.41**	**0.59**	421
	s_{30}	0.02	0.08	0.30	0	0.47	0.12	0.01	0.00	**0.40**	**0.60**	431
G_1	*basic*	0.00	0.01	0.09	0.52	0	0.33	0.04	0.01	**0.63**	**0.37**	484
	s_{10}	0.00	0.05	0.17	0.35	0	0.33	0.08	0.02	**0.57**	**0.43**	399
	s_{20}	0.00	0.05	0.14	0.39	0	0.33	0.09	0.00	**0.57**	**0.43**	422
	s_{30}	0.01	0.03	0.13	0.39	0	0.36	0.06	0.01	**0.56**	**0.44**	432
G_2	*basic*	0.00	0.01	0.12	0.54	0	0	0.25	0.07	**0.67**	**0.33**	263
	s_{10}	0.00	0.02	0.14	0.40	0	0	0.33	0.11	**0.56**	**0.44**	316
	s_{20}	0.00	0.02	0.13	0.44	0	0	0.30	0.10	**0.59**	**0.41**	293
	s_{30}	0.00	0.03	0.14	0.46	0	0	0.30	0.08	**0.62**	**0.38**	295
G_3	*basic*	0.01	0.02	0.10	0.52	0	0	0	0.35	**0.65**	**0.35**	102
	s_{10}	0.00	0.02	0.15	0.40	0	0	0	0.43	**0.57**	**0.43**	164
	s_{20}	0.00	0.02	0.16	0.46	0	0	0	0.35	**0.65**	**0.35**	161
	s_{30}	0.00	0.03	0.15	0.45	0	0	0	0.37	**0.63**	**0.37**	151
G_4	*basic*	0.03	0.03	0.38	0.56	0	0	0	0	**1.00**	**0.00**	66
	s_{10}	0.00	0.03	0.33	0.65	0	0	0	0	**1.00**	**0.00**	113
	s_{20}	0.00	0.04	0.38	0.58	0	0	0	0	**1.00**	**0.00**	92
	s_{30}	0.00	0.09	0.24	0.67	0	0	0	0	**1.00**	**0.00**	90

Table 2 Transition probabilities

As is obvious from the results in Table 2, a greater probability (certainty) can be seen in the model basic where the values range around 0.66, while with adaptive models they move around the value of 0.60. Adaptive models also occur more often with a bigger growth (G_3, G_4) and a bigger decrease (D_3, D_4).

Further on, the achieved value of capital for the monitored period is calculated. The value of capital is stated in Table 3. For each trading strategy the value of capital is C_n and the number of trades implemented is *n*. Trad-

ing strategies with varying state space which exceeded the strategy appropriate for them in the basic model are in bold.

		G_2		G_3		G_4	
		C_n	n	C_n	n	C_n	n
D_2	*basic*	1.800	173	1.646	91	1.200	55
	s_{10}	1.534	218	1.004	130	**1.211**	99
	s_{20}	1.265	201	1.502	133	**1.201**	79
	s_{30}	1.402	191	**1.692**	115	**1.264**	75
D_3	*basic*	1.935	76	0.939	54	0.991	40
	s_{10}	1.566	128	**1.043**	93	0.773	70
	s_{20}	1.910	119	**1.726**	90	**1.432**	66
	s_{30}	1.320	115	**1.202**	93	**1.201**	64
D_4	*basic*	1.966	50	1.582	39	1.481	34
	s_{10}	**3.569**	90	**3.872**	76	**1.497**	58
	s_{20}	**2.516**	82	**1.901**	73	1.374	51
	s_{30}	1.308	71	**1.832**	63	1.075	47

Table 3 Results of models

As is obvious from the results, each model with varying state exceeded the basic model in 5 trading strategies. Models with varying state space were successful in trading strategies D_2-G_4, D_3-G_3 and D_4-G_3. Models with varying state space have a bigger number of implemented transactions. The number of transactions grows with the length of the moving standard deviation. The average value of capital, calculated according to the equation (4) of the individual models (listed in the following order *basic*, s_{10}, s_{20}, s_{30}) is 1.505, 1.785, 1.643, 1.366. The development of the average value of capital can be seen in Figure 1. The graph shows, apart from the monitored models, also the development of B&H strategy with the reinvested dividends (0.869).

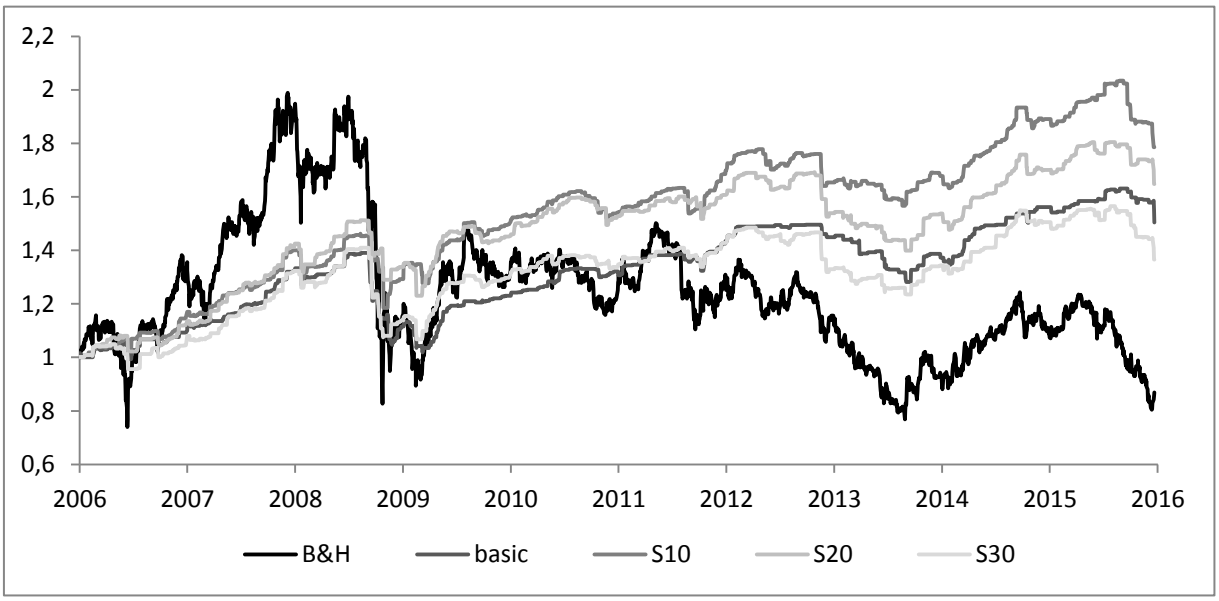


Figure 1 Development of average capital value

As is obvious from Figure 1, all models were achieving similar yields and were losing on B&H strategy up until the first half of year 2008. On the contrary, in the period of the dramatic fall of the share price in the second half of year 2008 and the beginning of year 2009 the models often showed smaller drops than B&H strategy. In that period the best results were achieved by models s_{10} and s_{20}. Since year 2010 the share price has showed periods of gradual falls and growths and the models were performing in a similar way but all of them outperformed B&H strategy. In Table 4 annual yields of the individual models are calculated.

Year	B&H	*basic*	s_{10}	s_{20}	s_{30}
2015	-21.6%	-3.6%	-5.5%	-3.1%	-9.3%
2014	19.4%	12.6%	11.9%	10.8%	12.3%
2013	-19.7%	-4.3%	2.1%	-0.3%	0.8%
2012	-9.6%	1.7%	-1.8%	-4.3%	-6.8%
2011	4.5%	7.9%	9.6%	5.7%	6.7%
2010	-5.2%	7.4%	2.7%	4.8%	3.1%
2009	14.6%	10.3%	15.8%	8.5%	14.0%
2008	-40.7%	-15.7%	-7.7%	-5.3%	-13.9%
2007	45.2%	21.0%	21.6%	25.8%	26.3%
2006	30.8%	9.3%	15.2%	12.3%	4.6%

Table 4 Annual returns of models

In the monitored ten year long period the model s_{10} outperformed the model basic in six instances out of ten, model s_{20} outperformed the model basic in five instances and model s30 outperformed the model basic only in three instances. In comparison with B&H strategy the models show, besides higher efficiency, also lower volatility.

4 Conclusion

In this study we focused on creating a technical indicator which uses the Markov chains analysis with varying state space for generating trading signals. The state space is defined parametrically as a multiple of the moving standard deviation. We calculated three models, one for the moving standard deviation with length 10, another with length 20 and yet another with length 30. The results of the study show that the trading strategies defined on these models generally outperform the passive Buy and Hold strategy. The best result is achieved by the model with the moving standard deviation with length 10. It can be stated that the shorter the length, the better the results. Models with lengths 10 and 20 outperformed the reference model with the unvarying state space. The differences were not big, though.

This initial study showed that the trading strategies applying the Markov chains analysis with varying state space can be profitable and that is the reason why we are going to continue the research even in the future. We intend to concentrate on the following areas:

- confirmation of the results of this study by applying them to other shares;
- a detailed analysis of the individual trading strategies, the ratio of profitable transactions, the average yield stemming from profitable trading transactions;
- a more detailed analysis of successfulness of the trading strategies based on growing, falling and lateral primary trends;
- modification of trading strategies.

Acknowledgements

The research project was supported by Grant SGS-2016-057 of the Faculty of Economics, University of West Bohemia in Pilsen

References

[1] Brock, W., Lakonishok, J., & LeBaron, B.: Simple Technical Trading Rules and The Stochastic Properties of Stock Returns. *Journal of Finance*, vol. 47 (1992), issue 5, 1731 – 1764.
[2] Doubleday, K., J., & Esunge, J., N.: Application of Markov Chains to Stock Trends. *Journal of Mathematics and Statistics*, vol. 7 (2011), issue 2, 103-106.
[3] Fama, E.: The Behavior of Stock Market Prices, *The Journal of Business*, vol. 38 (1965), issue 1, 34 -105.
[4] Hillier, F. S., Lieberman, G. J.: *Introduction to Operations Research*, 4th ed. McGraw-Hill Publishing Company. New York 1986
[5] Murphy, J.: *Technical analysis of the Financial markets*, New York: Penguin Group, 1999.
[6] Svoboda, M.: Stochastic Model of Short-Term Prediction of Stock Prices and its Profitability in the Czech Stock Market. *E&M Economics and Management*, vol.(2016), issue 2, 188-200.
[7] Sweeny, R., J.: Some new filter rule tests: methods and results. *Journal of Financial and Quantitative Analysis*, vol. 23 (1988), issue 3, 285–300.
[8] Vasanthi, S., Subha M., V., & Nambi S., T.: An empirical Study on Stock Index trend prediction using Markov Chain analysis. *JBFSIR*, vol.(2011), 72 – 90.

Long-Run Elasticity of Substitution in Slovak Economy: The Low-Frequency Suply System Model

Karol Szomolányi, Martin Lukáčik, Adriana Lukáčiková[1]

Abstract. New dataset provided by National Bank of Slovakia since 2014 includes the capital stock time series and enlarges possibilities to estimate the elasticity of substitution coefficient in the Slovak economy. The Actual approaches of production function parameters estimation use the data modification through low-pass filters to eliminate the effects of different underlying economic processes and the supply side system of the first-order conditions of a firm maximizing its profit. The studies keep the non-unitary substitution elasticity in the analysed economies.
In the paper, we estimate substitution elasticity coefficient in the Slovak economy using both approaches. The ordinary least square, seemingly unrelated regression and generalized method of moments is used. Data are modified by the low-pass filter procedure. In result we state that the Slovak elasticity of substitution is relatively small, seemingly unrelated regression methods provide even smaller estimation of the coefficient and that the estimated value of the coefficient is increasing with the periodicity selected in the low-pass filter process.

Keywords: elasticity of substitution, low-pass filter, supply side system

JEL Classification: C32, C36, E23, E25
AMS Classification: 91G70

1 Introduction

Chirinko [3] and Klump et al. [9] highlight the importance of the elasticity of input substitution with the rich literature review. Klump et al. [8] estimated the system of equations consisting of the constant elasticity of substitution production function linearization and two first-order conditions of a firm maximizing its profit. Authors accentuate the production function normalization and a biased technological change concept. Chirinko et al. [4] propose an approach of the coefficient of the elasticity of the substitution estimation dealing with the first-order condition of the firm maximizing its profit corresponding to the capital. Authors modify their data series by the low-pass filter to abstract them from the business cycles and the short-term effects driven by different underlying processes. They prove that the approach meets the production function normalization and they observe that considering a biased technological change does not affect the result elasticity of substitution estimate. Finally they observe that the method can be used for aggregate data. Both studies estimate low value of the elasticity of substitution in the U.S. economy: 0.60 by Klump et al. [8] and 0.40 by Chirinko et al. [4]. The literature review of the past empirical studies estimating the production function coefficients provide Chirinko [3] or Klump et al. [9]. In the paper we use the compromise approach of both Chirinko et al. [4] and Klump et al. [8] to estimate elasticity of substitution in Slovak economy. We use the low-pass filter to measure variables by the proper long-run values as Chirinko et al. [4] and we estimate the coefficients of system of first-order equations of a firm maximizing its profit.

2 Data and Methodology

The system of first-order conditions of a firm maximizing its profit equation system can be written as follows:

$$\begin{aligned} k_t - q_t &= \beta_{01} + \beta_1 \left(p_t^k - p_t^q \right) + \beta_{21} t + u_t \\ l_t - q_t &= \beta_{02} + \beta_1 \left(p_t^l - p_t^q \right) + \beta_{21} t + v_t \end{aligned} \tag{1}$$

where left-hand side of the system corresponds to the (log of) capital/output or labor/output ratio, the independent variable on the right-hand side corresponds to their relative prices. To eliminate the effects of different underlying economic processes, ensuring the exogeneity of the price ratios, both variables are measured by the proper long-run values. As Chirinko et al. [4] argue, these values can be reached using low-pass filter. The elasticity of substitution is the negative value of the β_1 coefficient.

[1] University of Economics in Bratislava, Dolnozemská cesta 1, Bratislava, Slovakia, karol.szomolanyi@euba.sk.

2.1 Data

Electronic macroeconomic database of the National Bank of Slovakia [11] provides the quarterly capital stock data series estimate. The data series are in current prices and seasonally adjusted. Therefore we use this portal to gather all other quarterly data series[2]. All gathered data are seasonally adjusted except interest rates. To assure the correctness of our process, we normalized all used price deflators to the value 1 in the 2010:Q4 period and re-computed the data-series in constant prices. We use real seasonally adjusted *GDP* and its deflator to measure the output volume and its price.

The challenge is to find the volume of capital, while the dataset of the capital stock is in current prices. For this purpose, we firstly need to compute the net investment deflator. The net investment in current prices is measured as the first difference of the capital stock. Let us define a net capital creation as the difference between gross capital creation and the consumption of the capital. We computed the net capital creation in both current and constant prices and so we gathered the net capital formation deflator. The deflator we use to compute net investment in constant prices from the nominal net investment. As our deflators are normalized for the 2010:Q4 period, the capital stock in constant prices equals to the capital stock in current prices in this period. Subtracting the net investment from capital stock in constant prices, we gain the capital stock in constant prices before this period, and, reversely, cumulating the net investment to the capital stock in constant prices, we gain capital stock in constant prices after this period.

The capital price is the sum of interest and depreciation rates. We use the interest rate on loans to non-financial corporations and interest rate on deposits of non-financial corporations. According to the two interest rate use, the data set version doubles. We measure the depreciation rate as the consumption-of-the-fixed-capital/capital-stock ratio in constant prices.

The labor can be measured by three official data series of employment gathered by Statistical Office of the Slovak Republic. The first data series is gathered using Labor Force Review and it measures number of persons. The second and third are gathered using European system of national and regional accounts (*ESA*). The second measures the number of persons and the third the hours worked. Since we have three measures of the labor, the data set versions are three times more. The indication of the dataset versions is in the Table 1.

dataset version	capital price measurement	labor measurement
(*a*)	loans to non-financial corporations	Labor Force Review
(*b*)	deposits of non-financial corporations	Labor Force Review
(*c*)	loans to non-financial corporations	*ESA* / persons
(*d*)	deposits of non-financial corporations	*ESA* / persons
(*e*)	loans to non-financial corporations	*ESA* / hours
(*f*)	deposits of non-financial corporations	*ESA* / hours

Table 1 Dataset versions

To achieve the labor price, we gather the labor income data series. Dividing by the labor, we gain the labor price. To derive dataset we followed Gollin [5] and Klump et al [8].Gollin [5] refers an inconsistency between a theory and observed values of labor share. This inconsistency comes from incorrect calculation of labor share. Compensation to employees is not suitable indicator for labor income because they exclude proprietors (self-employed) labor income. It is unclear how the income of self-employed workers should be categorized in the labor-capital dichotomy. Gollin [5] introduced several ways to modify data for correct labor share calculation. Our choice is limited by the published data. Some ways need indicators that are published but with small range, or they are annual, or we miss corresponding deflators. We use compensation per employee as a shadow price of labor of self-employed workers, i.e. labor income in extensive form is:

$$labour\ income = \left(1 + \frac{self\ employed}{total\ employment}\right) \cdot compensation\ to\ employees \tag{2}$$

Baxter and King [2] suggested appropriate low-pass filter. The filter is sensitive to the selection of frequency length (lead/lags) for the moving average, and the low and high values for the cycle period. We use the frequency respond function to find that Baxter and King's suggested selection of 3 year frequency length *m* (12 quarters), 1.5 year (6 quarters) low and 8 year (32 quarters) high cycle period is proper. We made experiments filtering data with other combinations of the frequency length and cycle periodicity. Our dataset uses the period 1997 – 2014 (2nd quarter)[3]. After filtering we lose the first *m* and the last *m* observations.

[2] http://www.nbs.sk/en/monetary-policy/macroeconomic-database/macroeconomic-database-chart
[3] The upper limit is given by an availability of the consumption of fixed capital in current prices data series.

2.2 Methodology

Under exogeneity assumption on the price ratio, the ordinary least square method is suitable to estimate the form (1). However, these two relations are probably related. Therefore the seemingly unrelated regression method is used as well. Moreover the supply and demand system of equations can be simultaneous. Since the low-pass filter creates overlapping observations, the stochastic term is serially correlated. Therefore we also estimated the system (1) by the generalized methods of moments and we compute standard errors with the procedure of Newey and West [12].We use the instrument set including constant trend and two lags of all variables. The unit root of residuals is tested by the augmented Dickey and Fuller and Phillips Peron tests (see Lukáčik et al. [10]) stating that residuals are stationary. We prefer the later test, if many autoregressive terms solved autocorrelation in the test specification.

3 Results

Table 2 provides the elasticity of substitution estimations using different datasets and different methods. The table rows correspond to dataset versions denoted in the Table 1. The table columns correspond to the method used: ordinary least square (**OLS**), seemingly unrelated regressions (**SUR**) and generalized method of moments (**GMM**). The estimations of corresponding elasticity of substitution coefficients are in the table. The corresponding standard deviation is in brackets below the each coefficient estimation. Using **GMM** estimation method, the standard deviations are computed with the procedure of Newey and West [12] (last column brackets). From these values it follows that all presented estimated coefficients are statistically significant at the 1% level.

dataset/method	OLS	SUR	GMM
(*a*)	0.142 [0.020]	0.093 [0.016]	0.146 [0.011]
(*b*)	0.084 [0.010]	0.060 [0.008]	0.083 [0.004]
(*c*)	0.144 [0.020]	0.087 [0.017]	0.143 [0.011]
(*d*)	0.084 [0.011]	0.057 [0.008]	0.083 [0.004]
(*e*)	0.150 [0.021]	0.109 [0.019]	0.150 [0.008]
(*f*)	0.086 [0.011]	0.069 [0.009]	0.086 [0.004]

Table 2 Elasticity of substitution estimations

The elasticity of substitution estimations do not depend on the labor data series choice. However, using the interest rate on deposits of non-financial corporations in the capital income calculation (*b*) (*d*) (*f*) decreases the estimated value of the coefficient as well as using the **SUR** method. We consider the interest rate on loans to non-financial corporations to be a proper measure.

4 Conclusion

Comparing our results with the novel estimates around the world, the Slovak elasticity of substitution is relatively low. However, the reviews of other empirical papers provided by Chirinko [3] and Klump et al. [9] admit such low values. We estimated the system (1) using different combinations of the frequency length and cycle periodicity. As well as Chirinko et al [4], we state that the estimated value of the elasticity substitution is increasing with the periodicity. However, the estimated values are below the estimates of mentioned novel papers.

Jürgen [6] provides a possible explanation of low Slovak elasticity of substitution. Using a theoretical model with micro-foundations he assumes a lower elasticity of substitution in transition economy. We assume that Slovakia has been the transition state in the study period. Using neoclassical growth conceptual schema, all events like the transition from central planned economy to the market economy, opening to the western European markets, entrancing to the European monetary Union, labor tax reforms have changed the Slovak steady state and have started Slovak transition move to its new steady state. In fact, using β-convergence estimation form of Barro et al. [1], Szomolányi et al. [13] and using spatial econometric tools Chocholatá et al. [7] observed that Slovakia has been in the transition state in the study period.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic - VEGA, grant no. 1/0444/15 "Econometric Analysis of Production Possibilities of the Economy and the Labour Market in Slovakia".

References

[1] Barro, R.J., and Sala-I-Martin, X.: *Economic Growth*. 2nd ed. The MIT Press, Cambridge, Massachusetts, London, England, 2003.
[2] Baxter, M. and King, R. G.: Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *The Review of Economics and Statistics* **81** (1999), 575-593.
[3] Chirinko, R.S.: σ: The Long and Short of It. *Journal of Macroeconomics* **30** (2008), June, 671–686.
[4] Chirinko, R. S., Mallick, D.: The Substitution Elasticity, Factor Shares, Long-Run Growth, and the Low-Frequency Panel Model. *CESifo Working Paper Series* 4895, CESifo Group Munich, July 2014 (this version December 2015).
[5] Gollin, D.: Getting Income Shares Right. *Journal of Political Economy* **110** (2), 2002, 458–474.
[6] Jürgen, A.: A Dual Elasticity of Substitution Production Function with an Application to Cross Country Inequality. *Economics Letters* **102** (2009), January, 10–12.
[7] Chocholatá, M., and Furková, A.: Does the location and institutional background matter in convergence modelling of the EU regions? *Central European Journal of Operations Research* **24** (2016), 1–19.
[8] Klump, R., McAdam, P., and Willman, A.: Factor Substitution and Factor Augmenting Technical Progress in the US. *Review of Economics and Statistics* **89** (2007), 183–192.
[9] Klump, R., McAdam, P., and Willman, A.: The Normalized CES Production Function: Theory and Empirics. *Journal of Economic Surveys* **26** (2012), 769–799.
[10] Lukáčik, M. and Lukáčiková, A.: Význam testovania stacionarity v ekonometrii. *Ekonomika a informatika*, **6** (2008), 146-157.
[11] National Bank of Slovakia: Macroeconomic database. Retrieved March 3, 2017, from http://www.nbs.sk/en/monetarypolicy/macroeconomic-database/macroeconomic-database-chart.
[12] Newey, W. K. and West K. D.: A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55 (1987), 703–708.
[13] Szomolányi, K., Lukáčiková, A. and Lukáčik, M.: The β-convergence of the EU27 Countries. In: *Proceeding of 29th International Conference on Mathematical Methods in Economics 2011*, (Dlouhy, M. and Skocdopolova, V., eds), Professional Publishing, Prague, Czech Republic, 2011, 287-291.

Stability analysis of optimal mean-variance portfolio due to covariance estimation

Blanka Šedivá [1], Patrice Marek [2]

Abstract. The objective of this paper is to study the stability of the mean-variance portfolio optimization. The results of the mean–variance optimal selection problem are very sensitive to the model parameters (portfolio calibration window and frequency of portfolio rebalancing). There are presented three methods of stabilization of covariance matrix estimation and further analysis are focused on the influence of estimation of covariance matrix to robustness of optimal selection.
For practical verification, the proposed approaches were tested the suitability of these methods to the performance of the investment portfolio. This were done within the framework of 2000 to 2016 using daily data of 100 companies from the New York Stock Exchange.

Keywords: Portfolio Selection Problem, Covariance Estimation, Mean-Variance Optimization; Random Matrix.

JEL classification: G11
AMS classification: 91G10

1 Introduction

There are many articles about optimal portfolio, the reason is that investing on a stock market with potential of profit is interesting for a large amount of people in the world. Thanks to this fact, there is also many approaches how mathematicians try to model the stock market. Some of them tend to believe that there is no relationship between the history and the future. This group use methods of random walk and tries to simulate large number of scenarios to forecast the future. The other group of scientists believe that there is a strong relation between historical prices and the future ones. But the determination the length of historical horizon is crucial problem in these methods.

The results of the mean-variance optimal selection problem are very sensitive to the model parameters (portfolio calibration window and frequency of portfolio rebalancing). There are presented three methods of stabilization of covariance matrix estimation and further analysis are focused on the influence of estimation of covariance matrix to robustness of optimal selection.

2 Covariance Estimation

Suppose we study a N- dimensional log price process $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$. This process corresponded with a set of financial assets which are characterized by their random log price in chosen time period, so the random vector is the vector of random log price on the individual assets. These prices are observed over a generic interval $[0, 1]$, which we think of as a day. The log price is modelled as $X(t) = \int_0^t a(u)du + \int_0^t \sigma(u)dW(u)$, where the elements of a are predictable locally bounded drifts, σ is a volatility process, and W is a vector of independent Brownian motions. Then, the ex-post covariation of X is given by $\mathrm{cov}(\boldsymbol{X}) = \int_0^t \Sigma(u)du$, where $\Sigma(u) = \sigma(u)\sigma^T(u)$.

The distribution of vector $\boldsymbol{X}$ is characterized by vector of expected values with elements $\mathbb{E}\, X_i$ and by covariance matrix Σ whose i, j^{th} element is the covariance between the X_i^{th} and the X_j^{th} random variables. The elements on diagonal σ_i^2 of matrix Σ represent variances of asset $i, i = 1, 2, \ldots, N$.

2.1 Basic Notation

If we want to measure and optimize the risk of the portfolio, it is necessary to use a reliable estimate of the covariance matrix Σ or correlation matrix $\boldsymbol{C}$. The estimation of correlation matrix has of the great importance for asset allocation but the difficulty in quantifying correlation between any two assets i, j arises from the following:

- There is no standardized method how to calculate the interaction "strength" between two companies i, j. Although there is an interaction between every pair of companies, but the precise nature of interaction is unknown.

[1]University of West Bohemia, Czech Republic, Department of Mathematics, Univerzitní 8, Pilsen, Czech Republic, sediva@kma.zcu.cz
[2]University of West Bohemia, Czech Republic, Department of Mathematics, Univerzitní 8, Pilsen, Czech Republic, patrke@kma.zcu.cz

- Correlations don't have to be just pairwise, but rather involving clusters of the assets.
- The correlations between any two pairs of stocks change with time and the cross correlation that exists between any pair of stocks may not be stationary.
- For each stock, we have only finite records of daily return of stock, from which we estimate an average correlation.

Suppose we have N stock the vector observations $\boldsymbol{X}(0), \boldsymbol{X}(1), \ldots, \boldsymbol{X}(T)$ for times $t_0, t_1, \ldots, t_T$ and we can defined the vector returns $\boldsymbol{r}(t) = (\boldsymbol{X}(t) - \boldsymbol{X}(t-1))/\boldsymbol{X}(t-1), t = 1, 2, \ldots, T$. It's mean that $r_i(t)$ is the daily return of stock i at time t, the empirical variance of each stock is given by

$$\sigma_i^2 = \frac{1}{T}\sum_t (r_i(t) - \overline{r}_i)^2 \tag{1}$$

and can be assumed for simplicity to be perfectly known. We also suppose, as usual, the daily return of stock $r_i(t)$ is demeaned ($\overline{r}_i = 0$). The empirical correlation matrix is obtained as

$$\widehat{\Sigma}_{ij} = \frac{1}{T}\sum_t x_i(t)x_j(t), \quad \text{where } x_i(t) = r_i(t)/\sigma_i \tag{2}$$

or in matrix form $\widehat{\boldsymbol{\Sigma}} = (1/T)\,\boldsymbol{X}^T\boldsymbol{X}$, where $\boldsymbol{X}$ is the normalization $N \times T$ matrix of the normalized return $X_{it} = r_i(t)/\sigma_i$.

2.2 Moving Average and Exponential Weighting Method

The simple setup for prediction models is use a random-walk forecasting model $\mathbb{E}(\Sigma(t+1)) = \widehat{\Sigma}(t)$, where $\widehat{\Sigma}(t)$ is the estimators of covariance matrix based on data for time $t - t_w$ to time t. The t_w is the length of historical estimation horizon for covariance matrix. It means that we will only use the previous data to perform asset allocation in the current period.

In practice applications a averaging of past may be used. This risks some bias but gains some estimation precision. The utilization of historical series of estimation of covariance matrix were be carried out in two ways: a h-moving average (denoted as **MA–h**) and an exponentially weighted moving average (denoted as $\mathbf{EWMA}_\alpha$).

The h-moving average method **MA–h** is defined by

$$\mathbb{E}_t(\Sigma(t+1)) = \left(\widehat{\Sigma}(t) + \widehat{\Sigma}(t-1) + \cdots + \widehat{\Sigma}(t-h+1)\right)/h \tag{3}$$

and in our empirical work we took $h = 5, 10$ and 20.

The exponentially weighted moving average method $\mathbf{EWMA}_\alpha$ is defined by

$$\mathbb{E}_t(\Sigma(t+1)) = (1-\alpha)\widehat{\Sigma}(t) + \alpha\,\mathbb{E}_{t-1}(\Sigma(t)) \tag{4}$$

and in our empirical work we took $\alpha = 0.5, 0.8$ and 0.9. We have chosen the values of the smoothing parameter that are recommended in [9].

2.3 Eigenvalue Cleaning Matrix

For a set of N different assets, the correlation matrix contains $N(N-1)/2$ entries, which must be determined from N time series of length T. If T is not very large compared to N, we can expect that the determination of covariances is noisy and therefore that the empirical correlation matrix is to a large extent random. Because a covariance matrix is positive semidefinite, that the structure of it can by describe by real eigenvalues and corresponding eigenvectors. Eigenvalues of the covariance matrix that are small (or even zero) correspond to portfolios of stocks that have non-zero returns, but extremely low or vanishing risk; such portfolios are invariably related to estimation errors resulting from insufficient data. One of the approaches used to eliminate the problem of small eigenvalues in the estimated covariance matrix is the so-called random matrix technique. Random matrix theory (RMT) was first developed by authors Dyson [5] and Mehta [11] for physical application, but there are also many results of interest in a financial context [8], [2], [13].

The spectral properties of $\boldsymbol{C}$ may be compared to those of random correlation matrix. As described by [8], [13] and others, if $\boldsymbol{R}$ is any matrix defined by $\boldsymbol{R} = (1/T)\,\boldsymbol{A}^T\boldsymbol{A}$, where $\boldsymbol{A}$ is an $N \times T$ matrix whose elements are i.i.d random variables with mean zero and fixed variance σ^2, than in the limit $T, N \to \infty$ keeping ratio $Q = T/N \geq 1$ constant, the density of eigenvalues of $\boldsymbol{R}$ is given by

$$P(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \quad \lambda_{min} \leq \lambda \leq \lambda_{max}, \tag{5}$$

where the maximum and minimum eigenvalues are given by

$$\lambda_{max/min} = \sigma^2 \left(1 \pm \sqrt{\frac{1}{Q}}\right)^2. \tag{6}$$

The distribution $P(\lambda)$ are known as the Marčenko-Pastur density [10] and the theoretical maximum and minimum eigenvalues determined the bounds for random matrix. If the eigenvalues of matrix are beyond these bounds, it is said that they deviate from random. If we apply this theoretical background of RMT to the correlation matrix we can separate the noise and non-noise parts of $\widehat{\mathbf{\Sigma}}$. We cleaned the matrix by following procedure:

- to construct the empirical correlation matrix as (2);
- to separate the noisy eigenvalues from non-noisy eigenvalues as (5);
- to keep the non-noisy eigenvalues the same and to take average of the noisy eigenvalues;
- to replace each eigenvalue associated with the noisy part by average of the eigenvalues;
- to reconstruct covariance matrix.

The simple repair mechanism, based on the spectral decomposition of the covariance matrix, is described for example in [7].

2.4 Composite Realized Kernel Cleaning Process

The other stabilization method of the estimation of covariance matrix is based on a composite realized kernels described in [1] or [9]. There is a data-efficient method, where the covariance estimate is composed of univariate realized kernels to estimate variances and bivariate realized kernels to estimate correlations. The basic principle of kernel method and its application to econometric models can be seen in [15]. The construction a multivariate realized kernel is given by following steps

- to construct the empirical correlation matrix $\widehat{\Sigma}(t)$ as (2) for every t;
- to calculate multivariate realized kernel as the form $RK(\boldsymbol{X}(t)) = \sum_{j=t-h/2}^{t+h/2} K\left(\frac{x}{h}\right) \widehat{\Sigma}(t)$, where $K(\cdot)$ is a kernel as a weighting function and h is a bandwidth. In our studies we used, following Barndorff-Nielsen et al. [1] and Lunge et al. [9], the Parsen kernel function in the form.

$$K(x) = \begin{cases} 1 - 6x^2 + 6x, & for \quad 0 \le x < 1/2 \\ 2(1-x)^3, & for \quad 1/2 \le x < 1 \\ 0, & for \quad x > 1 \end{cases} \tag{7}$$

  A main feature of multivariate kernels is that there is a single bandwidth parameter h which controls the number of leads and lags used for all the series.
- to calculate the "empirical kernel correlation" matrix $\widehat{\rho}_{ij}(t) = RK(X_i(t), X_j(t))/\sqrt{RK(X_i(t))RK(X_j(t))}$ and the optimal bandwidth h will be selected to suit the for each $\widehat{\rho}_{ij}$ i, j-th combination separately;
- to put elements RK and $\widehat{\rho}$ together and to obtain the "composite realized kernel matrix" as having ij-th element

$$CRK_{ij} = \widehat{\rho}_{ij}\sqrt{RK(X_i)RK(X_j)} \tag{8}$$

- Although RK is always positive semidefinite, the raw CRK matrix is not. So it is important for portfolio optimization to use impose positive semidefiniteness on CRK. We implement the regularization method used by Lunge et al. [9], which is known as eigenvalue cleaning or eigenvalue clipping.

3 Portfolio Construction and Statistical Performance

The covariance matrix estimates are key ingredients in portfolio optimizers. The Markowitz's theory of optimal portfolio [12] is focused on the problem to find optimal weight of each assets such that overall portfolio provides the best return for a fixed level of risk, or conversely the smallest risk for a given overall return [6].

3.1 Portfolio Theory - basic notations

More precisely the traditional Markowitz [12] portfolio problem is formulated as

$$\min \boldsymbol{w}' \mathbf{\Sigma}\, \boldsymbol{w}, \text{ subject to } \boldsymbol{w}'\boldsymbol{\imath} = 1, \quad \boldsymbol{w}'\boldsymbol{\mu} = \mu_0, \tag{9}$$

where $\boldsymbol{w}$ is the N by 1 vector of portfolio weights, $\mathbf{\Sigma}$ is the N by N covariance of returns, $\boldsymbol{\mu}$ is the N by 1 vector of expected returns, μ_0 is a target expected portfolio return and $\boldsymbol{\imath}$ is a conformable vector of 1's. The average return R_p of a portfolio P of N assets is defined as $R_p = \sum_{i=1}^{N} w_i r_i$ where w_i is the amount of capital invested in the

assets i and r_i are expected returns of the individual assets. Similarly the risk of a portfolio P can by associated with the total variance $\sigma_P^2 = \sum_{i,j=1}^{N} w_i \Sigma_{ij} w_j$ or in alternative form $\sigma_P^2 = \sum_{i,j=1}^{N} w_i \sigma_i C_{ij} \sigma_j w_j$ where σ_i^2 is the variance of asset i and C is the correlation matrix. The optimal portfolio which minimizes σ_P^2 for a given value of R_p can be easily found introducing a Lagrange multiplier and leads to a linear problem where the matrix $\boldsymbol{C}$ has to be invertible [4], [3].

The solution to the global minimum variance portfolio problem is

$$\boldsymbol{w}_C = \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\imath}\right) / \left(\boldsymbol{\imath}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\imath}\right). \tag{10}$$

3.2 Trading Turnover

In the previous section the several methods of estimating covariance matrix were by describe. If a reasonable estimate of covariance matrix is available, the structure of optimal portfolio,and we denoted the $w_j(t)$ the portfolio weights of asset j at time t. If we were to assume the cost of trading is proportional we can used for measurement of the cost of trading the method based on expression

$$c(t) = \sum_j |w_j(t) - w_j(t-1)|, \tag{11}$$

which is the cross-sectional average turnover of this method. The temporal average turnover for every method is defined as $\bar{c} = 1/T \sum_t c(t)$ and it is applied as one criterion for comparison of methods. The second criterion is concentrated on the results of the back-test performance of portfolio strategy that calculate cumulative value of one dollar invested at the beginning of time analysis.

4 Data Selection

We analysed assets prices for a large selection of the leading U.S. companies. As a prefered server was chosen server `http://chart.finance.yahoo.com`. For our analysis we used the time series of daily adjusted closing prices of stocks available on NYSE, which is the biggest stock market in the USA with over 3 147 assets at some time in the sample period January 2000 through December 2016. We restricted attention to exclude rather illiquid stocks in the sense that we focused on stock with the highest market capitalization. Our database Top100 includes 100 time series of the adjusted price that we followed over 4 352 trading days. The structure of all companies are in the following Table 1.

Sector	Number of available stocks	Percentage	Number in Top100	Percentage
Basic Industries	205	7%	7	7%
Capital Goods	183	6%	7	7%
Consumer Durables	58	2%	0	0%
Consumer Non-Durables	115	4%	10	10%
Consumer Services	448	14%	6	6%
Energy	226	7%	13	13%
Finance	351	11%	22	22%
Health Care	107	3%	17	17%
Miscellaneous	51	2%	4	4%
n/a	970	31%	0	0%
Public Utilities	208	7%	6	6%
Technology	160	5%	6	6%
Transportation	65	2%	2	2%
Total	3 147	100%	100	100%

Table 1 Structure of companies for our analysis

Throughout we worked with 100 time log of prices. Hence, the integrated variances of portfolio returns we reported in a moment are roughly the variance of percentage price daily moves of the portfolio.

As a representant of the market the NYSE index has been chosen and the risk free return has been calculated on the data of US Treasury Bonds Rates.

5 Empiric Results

The practical realization of described methods construction of optimal portfolio was performed with the following parameters: the historical estimation window for estimation raw covariance matrix $\widehat{\Sigma}$ was 120 trades days, the shift in time for each recalculation of optimal portfolio was 20 trades days.

For better understanding of the covariance structure of data we applied the method described in subsection 2.3. Firstly we constructed the empirical measured correlation matrix and we analysed the distribution of the eigenvalues. Basic summary statistics of eigenvalues are shown in the figure Fig. 2. Figure 2 clearly shows the results of our empiric studies of covariance matrix for period 2000-2016. Although the distribution of eigenvalues are similar for the whole period, the M-B bounder which depends on total variability cuts off only important covariances. The important information about asset mutual connections is carried in mean 23.2% of eigenvalues of the correlation matrix but there is a really big differences in time. Our previous simulation analyses show that this value is high but not unexpected [14].

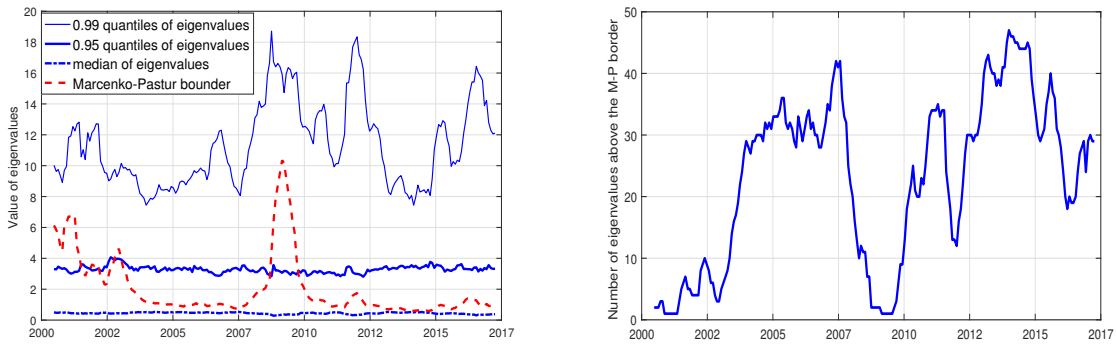


Figure 1 Basic quantiles of eigenvalues and Marcenko-Pastur bounder (left) and number of eigenvalues above the M-P border (right)

The results of tested methods are summarized in Table 2. For comparison, here are listed the cumulative value of investment for back-test Period from 23-Jun-2000 to 04-Apr-2017 and mean trading turnover.

Method	Parameters	Cumulative value of investment	Mean trading turnover
Cash		1.0817	0
Market		1.2292	0
MA-h	h=5	1.3315	0.8312
	h=10	1.3518	0.8447
	h=20	1.3035	0.8829
EWMA	α=0.5	1.4814	0.7514
	α=0.8	1.2657	0.7019
	α=0.9	1.2743	0.6882
Eigenvalue Cleaning Matrix		1.4898	0.7157
Composite Realized Kernel		**1.4928**	**0.7216**

Table 2 Results of the analysed methods for stabilization of estimation covariance structure.

6 Conclusion

The goal of this article was to develop an advance approach for stability treatment of portfolio optimization. There were presented three methods of stabilization of covariance matrix estimation and all this methods were by applied to the real data from NYSE within the framework of 2000 to 2016 using daily data of one hundred companies.

Our comparative analysis shown the advantages and disadvantages of several covariance estimator. In all cases, the results obtained were better than the results of the strategy portfolio strategy buy and hold. We found that the best performing forecasting model for our dataset is the composite realized kernel. The methods based on eigenvalue cleaning matrix provided comparable results.

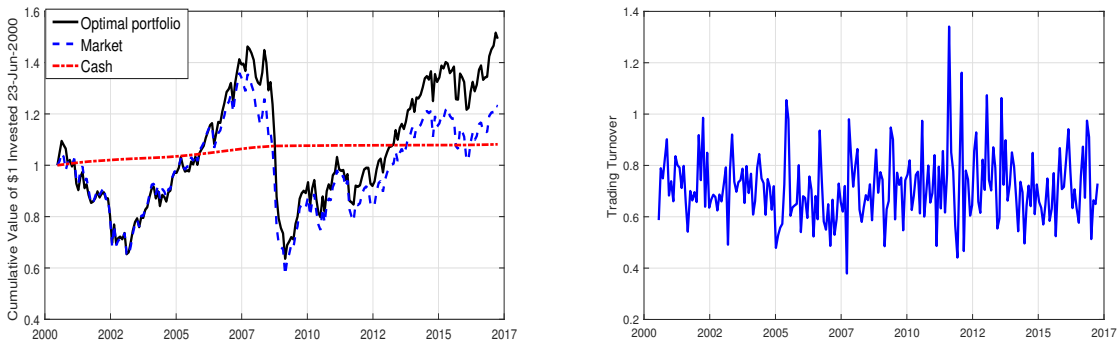


Figure 2 The result of back-test performance of the portfolio strategy (left) and trading turnover (right) for the composite realized kernel method

Acknowledgements

This publication was supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports.

References

[1] Barndorff-Nielsen, O. E., and Shephard, N.: Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society*, Series B, 64, pp. 253280.

[2] Daly, J., Crane, M., & Ruskin, H. J.: Random matrix theory filters in portfolio optimisation: a stability and risk assessment. *Physica A: Statistical Mechanics and its Applications* 387(16), (2008), pp. 4248–4260.

[3] Danielsson, J.: *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab* (Vol. 588). John Wiley & Sons, 2011.

[4] Dupačová, J.: Markowitzuv model optimální volby portfolia - předpoklady, data, alternativy. *(n.d.)*. Retrieved April 25, 2016,
from http://msekce.karlin.mff.cuni.cz/ dupacova/downloads/Markowitz.pdf.

[5] Dyson, Freeman J.: Correlations between eigenvalues of a random matrix. *Comm. Math. Phys.* 19, no. 3 (1970), pp. 235–250.

[6] Elton, E. J., Gruber, M. J., Brown, S. J. and Goetzmann, W. N.: *Modern portfolio theory and investment analysis.* John Wiley & Sons, 2009.

[7] Gilli, M., Maringer, D. and Schumann, E.: *Numerical methods and optimization in finance.* Academic Press, 2011.

[8] Laloux, L., Cizeau, P., Potters, M. and Bouchaud, J. P.: Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 3.03 (2000), pp. 391–397.

[9] Lunde, A., Shephard N. and Sheppard K.: Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business and Economic Statistics.* 2016, 34(4). pp. 504-518.

[10] Marčenko, V. A. and Pastur, L. A.: Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1.4 (1967), 457.

[11] Mehta, M. L.: *Random matrices* (Vol. 142). Academic Press, 2004.

[12] Markowitz, H.: Portfolio Selection *The Journal of Finance*, Vol. 7, No. 1 (Mar., 1952), pp. 77–91.

[13] Potters, M., Bouchaud, J. P. and Laloux, L.: Financial applications of random matrix theory: Old laces and new pieces. *arXiv preprint physics/0507111* (2005).

[14] Šedivá, B. & and Pavelec, J.: Adaptive parameter estimations of Markowitz model for portfolio optimization. In: *Proceedings of 34th International Conference Mathematical Methods in Economics, MME2016*, Technical University of Liberec, 2016.

[15] Ťoupal, T., Marek, P.: Nonparametric Kernel Estimation and Its Practical Application. In: *Proceedings of 8th International Scientific Conference Managing and Modelling of Financial Risks, MMFR 2016*, Ostrava: VB-TU, 2016. p. 1008 – 1015.

Long steps in IPM and L_1-regression

Barbora Šicková[1], Ondřej Sokol[2]

Abstract. This work deals with Newton method of the interior point which is applied to finding an estimation of polynomial L_1 regression. The aim of this paper is to find new modifications of long step choice in Newton method to find a faster solution of L_1 estimations for large data sets. For this purpose, five modifications of step length and two ways of its application in the algorithm are designed. Designed modifications are based on full-Newton step algorithm for the self-dual model.
Results are obtained in two phases. In the first phase, variable inicialization values are same for all algorithms. In the second, initialization values are changed to find a faster solution. Comparison of algorithms is based on the number of iterations where it is necessary to compute Newton step. According to the simulation study, the best results give algorithms which modify the barrier update parameter in an adaptive way. Algorithms and obtained results are implemented and visualized in MatLab.

Keywords: Newton method, L_1 regression, long step.

JEL classification: C61
AMS classification: 49M15

1 Introduction

This paper deals with the problem of choosing an optimal length of the step using interior point methods in linear programming. As the number of variables in linear models increases, the need to solve the problem efficiently arises. The simplex method may not be the best option to deal with some of the linear problems as the number of vertices may increase exponentially with the number of variables. There are other approaches to compute the optimal solution, such as ellipsoid, barrier or interior point methods. We focus on the third one.

The application of interior points methods to linear problems is an important breakthrough in solution efficiency. As it is showed by [3], the interior point method can be characterised by polynomial complexity and in some cases, those methods are more useful than the simplex method. The paper motivation is to find new modifications of optimal (or at least "good") long step choice in Newton method. We particularly focus on the Long-Newton step algorithm and the case of estimation parameters of polynomial regression under the L_1 norm.

We are inspired by the paper [5], in which authors compares the efficiency of different IPMs for L_1-regression. An estimation of polynomial L_1 regression is a robust method. This is the reason why it is one of the best alternative to the least squares method [1]. However, unlike [5], algorithms in this paper are based on the Full-Newton step algorithm.

Usually, full-Newton step method is looking for an optimal solution in such way that in each iteration Newton step length is reduced. Our method is based on an adaptive estimation of the optimal length of the next step.

The article is organised as follows: In Sections 2 and 3, we formally define the polynomial regression with L_1 norm and the Long-Newton step algorithm. In Section 4 we introduce the modifications of step choice and in Section 5 we present the results of simulation experiments. The paper concludes with a short summary.

2 Polynomial regression with L_1 norm as an optimization problem

Assume an polynomial L_1 regression

$$y_i = \sum_{j=0}^{m} \beta_j x_i^j + \epsilon_i, \quad i = 1, \ldots, n \tag{1}$$

where y_i is dependent variable, x_i is explanatory variable, ϵ_i is a random error component, m is the degree of polynom we use to explain dependent variable and n is the number of observations. In the general linear regression

[1]University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, 13067 Prague 3, Czech Republic, xsicb00@vse.cz
[2]University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, 13067 Prague 3, Czech Republic, ondrej.sokol@vse.cz

model with L_1 norm, the goal is to find the best estimation of β such that

$$\sum_{i=1}^{n} |y_i - \sum_{j=0}^{m} b_j x_i^j| \tag{2}$$

where b is the estimator of β, is minimized.

The model can be transformed to linear optimization problem:

$$\min_{b,r} \sum_{i=1}^{n} r_i \tag{3}$$

subject to (4)

$$y_i - \sum_{j=0}^{m} b_j x_i^j \leq r_i \quad i = 1, \ldots, n \tag{5}$$

$$-y_i + \sum_{j=0}^{m} b_j x_i^j \leq r_i \quad i = 1, \ldots, n \tag{6}$$

This problem is indeed linear; however, with the increasing degree of polynomial $\sum_{j=0}^{m} b_j x_i^j$ we may encounter computational problems. One way to solve this problem is to use some of the interior-point methods (see for example [2] and [4]) – in this paper we use Long-Newton step algorithm. Our main goal is to propose a method to choosing the optimal length of the step to reduce the total number of iterations.

3 Long-Newton step algorithms

Assume

$$\min\{\boldsymbol{q}^T \boldsymbol{z} : \boldsymbol{M}\boldsymbol{z} \geq -\boldsymbol{q}, \boldsymbol{z} \geq \boldsymbol{0}\}. \tag{7}$$

to be self-dual embedding and $\bar{n}$ is a dimension of matrix $\boldsymbol{M}$.

Let have a positive number μ, a nonnegative vector $\boldsymbol{z}$ and its slack vector $\boldsymbol{s}(\boldsymbol{z}) = \boldsymbol{M}\boldsymbol{z} + \boldsymbol{q}$, such that:

$$\boldsymbol{z}\boldsymbol{s}(\boldsymbol{z}) = \mu \boldsymbol{e}, \quad \boldsymbol{z} \geq \boldsymbol{0}, \quad \boldsymbol{s}(\boldsymbol{z}) \geq \boldsymbol{0} \tag{8}$$

If $\mu > 0$, then there is at most one nonnegative vector $\boldsymbol{z}$ such, that (8) holds. A solution of (8) is $\boldsymbol{z}(\mu)$ so called μ-center. If $\mu = 1$ than the existence of $\boldsymbol{z}$ is guaranteed.

Denote $\Delta\boldsymbol{z}$ as a displacement in z-space and $\Delta\boldsymbol{s}$ as a displacement in s-space defined as $\Delta\boldsymbol{s} = \boldsymbol{M}\Delta\boldsymbol{z}$. We want to find $\Delta\boldsymbol{z}$ such that $\boldsymbol{z}^+ = \boldsymbol{z} + \Delta\boldsymbol{z}$ is μ-centrum. In the Newthon method we solve linear system:

$$\begin{aligned} &\boldsymbol{M}\Delta\boldsymbol{z} - \Delta\boldsymbol{s} = \boldsymbol{0}, \\ &\boldsymbol{z}\Delta\boldsymbol{s} + \boldsymbol{s}\Delta\boldsymbol{z} = \mu\boldsymbol{e} - \boldsymbol{z}\boldsymbol{s}. \end{aligned} \tag{9}$$

Solution $\Delta\boldsymbol{z}$ is so called Newton direction.

To quantify an estimation quality we use a proximity measure $\delta(\boldsymbol{z}, \mu)$. $\delta(\boldsymbol{z}, \mu)$ is defined as $\delta(\boldsymbol{z}, \mu) = \frac{1}{2}\|\boldsymbol{v} - \boldsymbol{v}^{-1}\|$, where a variance vector $\boldsymbol{v}$ is defined as $\boldsymbol{v} = \sqrt{\frac{\boldsymbol{z}\boldsymbol{s}(\boldsymbol{z})}{\mu}}$.

The Full-Newton step algorithm is based on a Newton step $\boldsymbol{z}^+$ sequence where we count step length in each iteration. At the begining of the algorithm there is an accuracy parameter $\varepsilon > 0$ and a barrier update parameter θ, $0 < \theta < 1$. The algorithm starts at the 1-center, where $\mu = 1$. We choose θ such that Newton method is quadratically convergent. Let $\theta = \frac{1}{\sqrt{2\bar{n}}}$ then at the begining of each iteration $\boldsymbol{q}^T\boldsymbol{z} = \bar{n}\mu$ and $\delta(\boldsymbol{z}, \mu) \leq \frac{1}{2}$. All definitions and notation are from book [7]. For the brief description of Newton method see Algorithm 1.

Data: $\varepsilon > 0$, $0 < \theta < 1$, $\boldsymbol{z} = \boldsymbol{e}$, $\mu = 1$
while $\bar{n}\mu \geq \varepsilon$ **do**
  $\mu = (1 - \theta)\mu$;
  $\boldsymbol{z}^+ = \boldsymbol{z} + \Delta\boldsymbol{z}(\boldsymbol{z}, \mu)$; // $\Delta\boldsymbol{z}(\boldsymbol{z}, \mu)$ *is solution from* (9)
end

Algorithm 1: Full-Newton step

3.1 Adaptive full-long step algorithms with TRANSTHETA(θ) and with TRANSP(p) transformation function

Algorithm 2 schematically describes adaptive full-long step algorithm with TRANSTHETA(θ) transformation. And Algorithm 3 schematically describes adaptive full-long step algorithm with TRANSP(p) transformation.

[1] Initialization;
Data: $\varepsilon = e^{-8}$; $\theta^+ = 0,5$; $\mathbf{z} = \mathbf{e}$; $\mu = 1$; $\bar{\mu} = \mu$
[2] Outside loop;
while $\bar{n}\mu \geq \varepsilon$ **do**
  $\theta = 0,5$; // *long step attempt guaranty*
  $\bar{\mu} = (1-\theta)\mu$;
  $\mathbf{z}^+ = \mathbf{z} + \Delta\mathbf{z}(\mathbf{z}, \bar{\mu})$; // *long step attempt*
  [3] Inside loop with adaptive transformation, used when step unsuccessful;
  while $\delta(z^+, \bar{\mu}) > \frac{1}{2}$ **do**
    if $\theta = 0,5$ **then**
      $\theta = \theta^+$; // *transformation*
      $\bar{\mu} = (1-\theta)\mu$;
      $\mathbf{z}^+ = \mathbf{z} + \Delta\mathbf{z}(\mathbf{z}, \bar{\mu})$; // *step attempt*
    end
    if $\delta(z^+, \bar{\mu}) > \frac{1}{2}$ **then**
      θ = TRANSTHETA(θ); // *unsuccessful step, transformsation*
      $\bar{\mu} = (1-\theta)\mu$;
      $\mathbf{z}^+ = \mathbf{z} + \Delta\mathbf{z}(\mathbf{z}, \bar{\mu})$; // *step attempt*
    end
  end
  $\mathbf{z} = \mathbf{z}^+$; // *succesful step*
  $\mu = \bar{\mu}$;
  $\theta^+ = \theta$;
end

Algorithm 2: Adaptive full-long step algorithm with TRANSTHETA(θ) transformation function

[1] Initialization;
Data: $\varepsilon = e^{-8}$; $p^+ = 0,01$; $\mathbf{z} = \mathbf{e}$; $\mu = 1$; $\bar{\mu} = \mu$
[2] Outside loop;
while $\bar{n}\mu \geq \varepsilon$ **do**
  $p = 0,01$; // *long step attempt guaranty*
  $\theta = \frac{1}{\sqrt{p\bar{n}}}$; // *count barrier update parameter*
  $\bar{\mu} = (1-\theta)\mu$;
  $\mathbf{z}^+ = \mathbf{z} + \Delta\mathbf{z}(\mathbf{z}, \bar{\mu})$; // *step attempt*
  [3] Inside loop with adaptive transformation, used when step unsuccessful;
  while $\delta(z^+, \bar{\mu}) > \frac{1}{2}$ **do**
    if p $= 0,01$ **then**
      $p^+ = p$; // *transformation*
      $\theta = \frac{1}{\sqrt{p\bar{n}}}$; // *count barrier update parameter*
      $\bar{\mu} = (1-\theta)\mu$;
      $\mathbf{z}^+ = \mathbf{z} + \Delta\mathbf{z}(\mathbf{z}, \bar{\mu})$; // *step attempt*
    end
    if $\delta(z^+, \bar{\mu}) > \frac{1}{2}$ **then**
      p = TRANSP(p); // *transformation*
      $\theta = \frac{1}{\sqrt{p\bar{n}}}$; // *count barrier update parameter*
      $\bar{\mu} = (1-\theta)\mu$;
      $\mathbf{z}^+ = \mathbf{z} + \Delta\mathbf{z}(\mathbf{z}, \bar{\mu})$; // *step attempt*
    end
  end
  $\mathbf{z} = \mathbf{z}^+$; // *succesful step*
  $\mu = \bar{\mu}$;
  $p^+ = p$;
end

Algorithm 3: Adaptive full-long step algorithm with TRANSP(p) transformation function

4 Modifications of long step choice

We use the barrier update parameter $\theta = \frac{1}{\sqrt{p\bar{n}}}$ to define Newton step length. For the Full-Newton step algorithm we have $p = 2$. Long-Newton step transformations are based on two principles. In the first case, we transform θ by so called TRANSTHETA(θ) function, see Algorithm 2. Secondly, TRANSP(p) function changes parameter value p and then we count θ with using transformed parameter, see Algorithm 3.

The parameter transformation is applied in the Newton step algorithm in an adaptive way. At the beginning of each iteration, the algorithm tries to use long step. If it is not possible, the barrier update parameter is transformed. During transformation it is necessary to check if $\delta(\boldsymbol{z}, \mu) \leq \frac{1}{2}$ to hold convergency.

We have five pairs of modifications designed. First modification uses transformation with TRANSTHETA(θ) function and second uses TRANSP(p). For transformation definitions see Table 1, where d is the degree of estimated polynomial. Numerical values in described modifications are used in the first phase of results obtaining. In the second phase, inicialization values p, θ (*init*) from algorithms and numerical values (*num*) from described modifications are optimized in order to minimize the total number of iterations.

TRANSTHETA	Transformation	TRANSP	Transformation
AF-L	$\theta - 0,001$	AF-LP	$p + 0,001$
AF-L1	$\theta - 0,001d$	AF-LP1	$p + 0,001d$
AF-L-FP	$\theta/2$	AF-LP-FS	$2p$
AF-L-Pdd	$\theta/(d+1)$	AF-LP-Sdd	$(d+1)p$
AF-L-mixed	$\theta - \theta/(d+1)$	AF-LP-mixed	$p + (d+1)p$

Table 1 Algorithm review

5 Experiments and results

Algorithms are compared by the number of iterations – recomputations of the Newton step length (*steps*). For each different algorithm calculation, we generate 500 data sets to be sure that outlying values do not affect results. The degree of an estimated polynomial varies from 1 to 5. Independent variable values are 500 equally spaced points from interval $[0, 1]$. Dependent variable values are randomly generated using normal distribution $N(0, 1)$. Definitions of independent and dependent variables are based on the papers [5], [6].

In the first phase, we use the same initialization values p, θ and numerical values from transformations, as those defined in the previous sections. In the second phase, initialization values p, θ and numerical values from transformations are modified in order to find a more efficient solution. Values are defined according to experimental observations. In order to easier results comparison, axes in graphs are fixed. Tables 2 and 3 show average *steps* number. We show only results obtained in the second phase because only these results are relevant. Results obtained in the first phases are used only as a guide to finding a more efficient solution in the second phase.

5.1 Adaptive full-long step algorithm with TRANSTHETA(θ) - reasults

The table and graphs below shows average number of iterations where it is necessary to compute Newton step (*steps*), see Table 2 and Graph 1. We display the results of the second phase, in which the inicialization values for θ (*init*) and numerical values (*num*) are optimized.

alg/*degree*	1	2	3	4	5	*init*	*num*
AF-L	82,46	82,94	83,96	93,22	138,41	0,4	0,05
AF-L1	120,44	98,63	91,65	95,58	129,78	0,35	0,003
AF-L-FP	89,31	90,81	92,32	97,39	140,71	0,5	1,1
AF-L-Pdd	85,10	89,50	110,53	155,68	248,38	0,4	0,5
AF-L-mixed	82,39	87,20	90,76	95,18	136,11	0,45	2

Table 2 Adaptive full-long step algorithm with TRANSTHETA(θ) (*steps*)

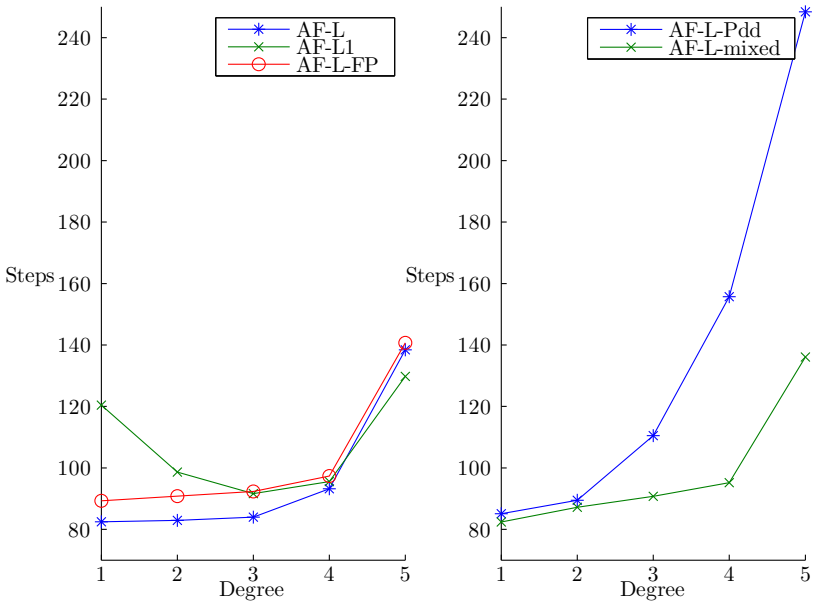


Figure 1 Adaptive full-long step algorithm with TRANSTHETA(θ) (*steps* evolution)

It is shown, that finding an algorithm which is more efficient depends on the estimated polynomial degree. Values *init* and *num* can vary according to number of independent variables. Algorithms AF-L, AF-L-FP and AF-L-mixed are better for estimation polynomial with degree between 1 and 3 while they perform approximately the same. For estimation of higher polynomial degree it is better to use AF-L1. On average, the most efficient are algorithms AF-L and AF-L-mixed.

5.2 Adaptive full-long step algorithm with TRANSP(p) - reasults

The table and graphs below shows average number of *steps*, see Table 3 and Graph 2. We again display the results of the second phase, in which inicialization values for p (*init*) and numerical values (*num*) are optimized.

alg/*degree*	1	2	3	4	5	*init*	*num*
AF-LP	77,35	77,63	78,18	87,17	123,09	0,005	0,01
AF-LP1	81,17	81,14	81	93,47	123,09	0,005	0,002
AF-LP-FS	83,64	84,59	86,17	93,08	136,15	0,003	1,5
AF-L-Sdd	87,19	90,91	93,03	93,03	136,15	0,0085	1
AF-LP-mixed	82,03	100,61	81,07	9636	119,69	0,004	0,1

Table 3 Adaptive full-long step algorithm with TRANSP(p) (*steps*)

We can see, that finding an efficient algorithm depends on estimated polynomial degree. Values *init* and *num* can vary according to number of independent variables. In this case algorithms AF-LP, AF-LP1 and AF-LP-FS are better for estimation polynomial with degree between 1 and 3 but for higher estimated polynomial degree too. They perform approximately same way as algorithms AF-L, AF-L-FP and AF-L-mixed with TRANSTHETA(θ) transformation. On average, the most efficient are algorithms AF-LP and AF-LP1.

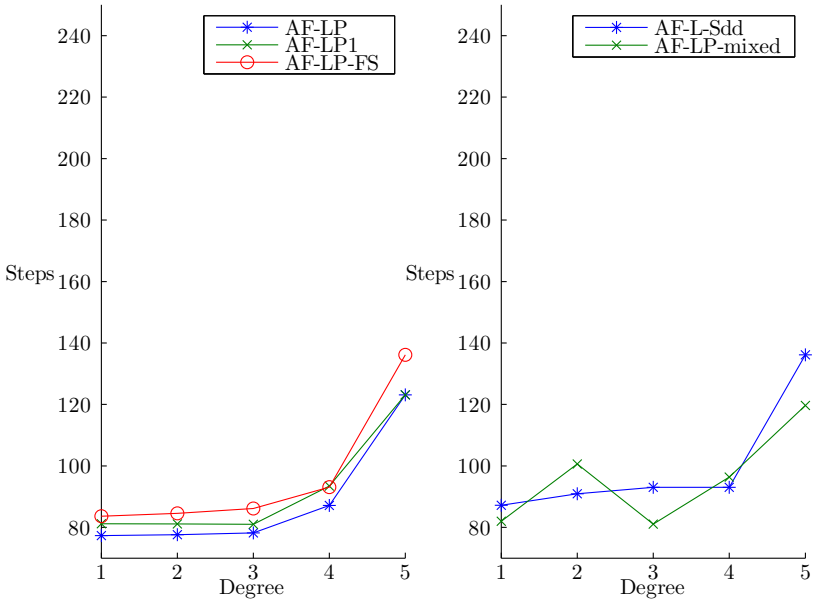


Figure 2 Adaptive full-long step algorithm with TRANSP(p) (*steps* evolution)

6 Conclusions

The aim of this paper is to design new modification of long step choice in order to find a faster solution of L_1 estimation. For this purpose, we separate experiments into two phases. In the first phase, variable inicialization values are same for all algorithms. In the second, initialization values are changed to find a faster solution. We compare algorithms according to the number of iterations in which it is necessary to compute new length of the Newton step. With designed modifications, we are able to find a solution of estimation of polynomial L_1 regression efficiently. According to the average number of iterations the best modifications were AF-L with TRANSTHETA(θ) transformation and AF-LP, AF-LP1 with TRANSP(p) transformation. These transformations are able to find L_1 estimation with less than average 97 Newton step recomputing iterations. In comparison with full Newton step, where average recomputing iterations number is 1404 it is a significant improvement.

Acknowledgements

This work was supported by grant No. 58/2017 of the Internal Grant Agency of the University of Economics, Prague.

References

[1] Dodge, Y.: Special issue on statistical data analysis based on the L_1 norm and related methods, *Computational Statistics & Data Analysis* **5** (1987), 237–4535.

[2] Freund, R. M., Mizuno, S. Interior Point Methods: Current Status and Future Directions. *In High Performance Optimization*, Boston, MA: Springer US, 2000.

[3] Karmarkar, N.: A new polynomial-time algorithm for linear programming, *Combinatorica* **4** (1984), 373–395.

[4] Potra, F. A., Wright, S. J. Interior-point methods. *Journal of Computational and Applied Mathematics* . **124** (2000), 281-–302.

[5] Oliveira, A. R. L. Nascimento, M. A. Lyraa, C.: Eficient implementation and benchmark of interior point methods for the polynomial L_1 fitting problem, *Computational Statistics & Data Analysis* **35** (2000), 119–135.

[6] Oliveira, A. R. L.Lyraa, C.: Interior point methods meet simplex in L_∞ fitting problems, *International Transactions in Operational Research* **115** (2004), 309–322.

[7] Roos, C. Terlaky, T. Vial, J.-P.: *Interior point methods for linear optimization*. New York: Springer Science + Business Media, Inc., 2005.

User versus Automatic Selection of Models in Actuarial Demographics: The Impact on the Expected Development of the Probability of Death in the Czech Republic

Ondřej Šimpach[1], Marie Pechrová[2]

Abstract. For the needs of actuarial demographics, it is not sufficient to construct only deterministic models, but it is suitable to combine the results also using stochastic ones. Finding optimal form of the model is a complicated process, because it is necessary to test not only statistical significance of the parameters, but also the significance of the whole model and to respect the results of the diagnostics tests. While applying of this modelling on data of age-and-sex specific mortality rates, there is a problem in the number of those models. There are about 2 × 101 or more models in advanced populations, as each gender is necessary to model separately and the age range is usually surveyed by statistical offices in the intervals 0–100+ completed years of life.

The aim of the paper is to assemble 2 × 101 optimal ARIMA models on the data of logarithms of age-and-sex specific mortality rates in the Czech Republic. Those models are diagnosed and consequently used for predictions. Other 2 × 101 ARIMA models are created on the same data, but using automatic process in software RStudio. It is applied a compromise form of the model on all time series. Both approaches - user and automatic are compared based on the results of the projection, that is recalculated on the probability of death of *x*-year old person. It was proved that automatic process is significantly faster and that the results of the projections are not distorted.

Keywords: ARIMA, mortality, forecast, probability of death, RStudio.

JEL Classification: C32, C55, J11
AMS Classification: 60G25

1 Introduction

"Mortality rates of human populations in developed countries are declining with time" (Finkelstein, [6]). With increasing life expectancy improving long-term care and sustaining the pension system are becoming an important issue. Besides, establishing methodologically sound longitudinal data sets is necessary in order to examine the phenomena (Andel [1]). For the purposes of actuarial demography, it is not sufficient to construct only deterministic models, but it is necessary to combine them and to compare the results also which stochastic ones. "The actuarial and demographic literature has introduced a myriad of (deterministic and stochastic) models to forecast mortality rates of single populations" (Antonion, Bardoutsos and Ouburg [2]). Work of Gompertz [10] played an important role in shaping the emerging statistical science. Gompertz model provided a powerful stimulus to examine the patterns of death ("law of mortality") across the life course not only in humans but also in a wide range of other organisms (Kirkwood [13]). Since that many of models have been developed. Lee and Carter [14] published a new statistical method for forecasting mortality in 1992. Since that it has been applied on many real data of the populations. For example, Li and Lee [15] applied the Lee-Carter model to a group of populations, allowing each its own age pattern and level of mortality but imposing shared rates of change by age.

However, there are more models used. For example, Antonion, Bardoutsos and Ouburg [2] presented in their paper a Bayesian analysis of two related multi-population mortality models of log-bilinear type, designed for two or more populations. Gogola [8] used stochastic mortality – Cairns, Blake and Dowd model that is well suited at very high ages to calculate mortality rates of age categories from 85 to 115 for selected countries. Godunov [9] compared several models applied on particular populations and found that „the most appropriate models of smoothing mortality curve are Kannistö-Thatcher (UK) Martinell (Sweden) and Kannistö (Canada). On the other side, the least suitable models are Coale-Kisker (Singapore), Gompertz-Makeham and modified Gompertz-Makeham (Czech Republic, Slovakia, Germany)". Mortality rates in the Czech Republic was calculated for example by Jindrová and Slavíček [12]. They applied Lee-Carter model on Czech population in the period of 1950–2009 and

[1] University of Economics Prague, Department of Statistics and Probability, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, ondrej.simpach@vse.cz.
[2] Institute of Agriculture Economics and Information, Mánesova 1453/75, 120 00 Prague 2, Czech Republic, pechrova.marie@uzei.cz.

predicted the development of specific mortality rates and consequently life expectancy for period 2010–2029. In the broader context, Fiala and Langhamrová [5] analysed of the development of the sex-and-age structure of the Czech population of productive age based on the latest population projection of the Czech Statistical Office. Probabilistic projections of age-specific mortality and fertility rates were done for example by Ševčíková et al. [16] in order to apply probabilistic population projections on United Nations (UN) countries.

Finding optimal form of the model age-and-sex specific mortality rates $m_{x,t}$ (respectively probability of death ($q_{x,t}$) after recalculation according to life table algorithm) is a complicated process, because it is necessary to test not only statistical significance of the parameters, but also the significance of the whole model and to respect the results of the diagnostics tests. Finding a suitable model for $m_{x,t}$ raises a problem that there are about 2 × 101 or more models in advance populations, because each gender and age[3] is necessary to model. Therefore, the aim of the paper is to compare user and automatic ARIMA approaches towards the model selection based on the results of the mortality projection that is recalculated on the probability of death of x-year old person. Paper shows that automatic process is significantly faster and that the results of the projections are not significantly distorted.

2 Methods and Data

Mortality is one of the demographic indicators showing the percentage of deaths for a certain period from a group of people. The age-specific mortality rates are one of the basic quantities used in modelling mortality. They are calculated according to the formula (1)

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}}, \tag{1}$$

where $m_{x,t}$ is age-specific mortality rate in age x and time t, $D_{x,t}$ is the number of deaths at completed age x in time t, and $E_{x,t}$ is mid-year state of x-year old population in time t, i.e. exposure to risk. Based on the observation of large population groups using demographic methods it is possible to estimate the probability of death for males and females of different ages and other important characteristics. From mortality tables at each year are calculated probabilities of death ($q_{x,t}$) at age x and time t as (2)

$$q_{x,t} = 1 - e^{-m_{x,t}}, \tag{2}$$

where e represents Euler's constant. Probabilities of survival ($p_{x,t}$) at age x and time t are therefore calculated as $p_{x,t} = 1 - q_{x,t}$. (Next steps can be seen, for example, in paper by Šimpach and Langhamrová [17]).

Data about of age-and-sex specific mortality rates for the period 1920–2015 in the Czech Republic were obtained from Czech Statistical Office (CZSO). They were consequently transformed to logarithms, tested by Dickey-Fuller test (ADF test) whether they were stationary and Box-Jenkins (Box and Jenkins [4]) methodology was applied on them. There are 3 types of ADF test with constant and trend, with constant only, and without constant and trend (i.e. without deterministic elements). The first case is calculated according the equation (3)

$$\Delta Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + \sum_{i=1}^{m} \alpha_i Y_{t-i} + \varepsilon_t, \tag{3}$$

where Δ is the first difference of the examined variable, t is the trend variable in this case, ε_t is pure white noise error term, m is the maximum length of the lagged dependent variable, and α, β are parameters (β_1 represents the constant). Box and Jenkins [4] introduced the models that are working with autoregressive (AR(p)) and moving average (MA(q)) processes. When the time series is not stationary, its difference of d^{th} order must be done. Diagnostic of the model type is done by Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) that were plotted in order to determine the order p of AR process and order q of MA process. General ARIMA(p,d,q) model is formulated as (4).

$$Y_t = \beta + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j} \tag{4}$$

There were assembled 2 × 101 optimal ARIMA models on the data of logarithms of age-and-sex specific mortality rates in the Czech Republic. From above stated information can be seen that wide diagnostic of the

[3] Age range is usually surveyed by statistical offices in the intervals 0–100+ completed years of life.

models is needed. Appling all tests to 202 time series manually can be time consuming. Therefore, other 2 × 101 ARIMA models were applied on the same data, but using automatic process in software RStudio. Both approaches user and automatic are compared based on the results of the projection, that is recalculated on the probability of death of *x*-year old person.

3 Results

From the empirical data of age-and-sex specific mortality rates in logarithms (see Fig. 1, top charts) is evident, that there was unstable development of time series of mortality rates until 1950 (especially in the case of male population). The age-and-sex specific mortality rates are lower in the case of female population, because, in general, the intensity of mortality ($\mu_{x,t}$ according to Gompertz law) of the female population is lower in most age groups. Significant difference between male and female mortality is in age group 18–32 years and at the oldest age groups (85+). Higher mortality level of young males is caused by suicides, poisoning, dangerous behaviour, gambling, etc. (this is unfortunately a long-term trend in most populations, not only in the Czech Republic). Recalculated values of age-and-sex specific probabilities of death according to the life table algorithm are shown in Fig. 1 (bottom charts). It is well known that the instability of the time series reduces their predictive capability (Bell [3] or Gardner, McKenzie [7]). The history, although, has the lowest weight in the prediction model. For modelling of mortality, which is a long-term process that has for each population its long-term trend, the history (even with a little weight) could be quite important. We do not smooth the empirical data in this paper according to Gogola [8] or Godunov [9] and rather use the ARIMA models with or without constant on raw data.

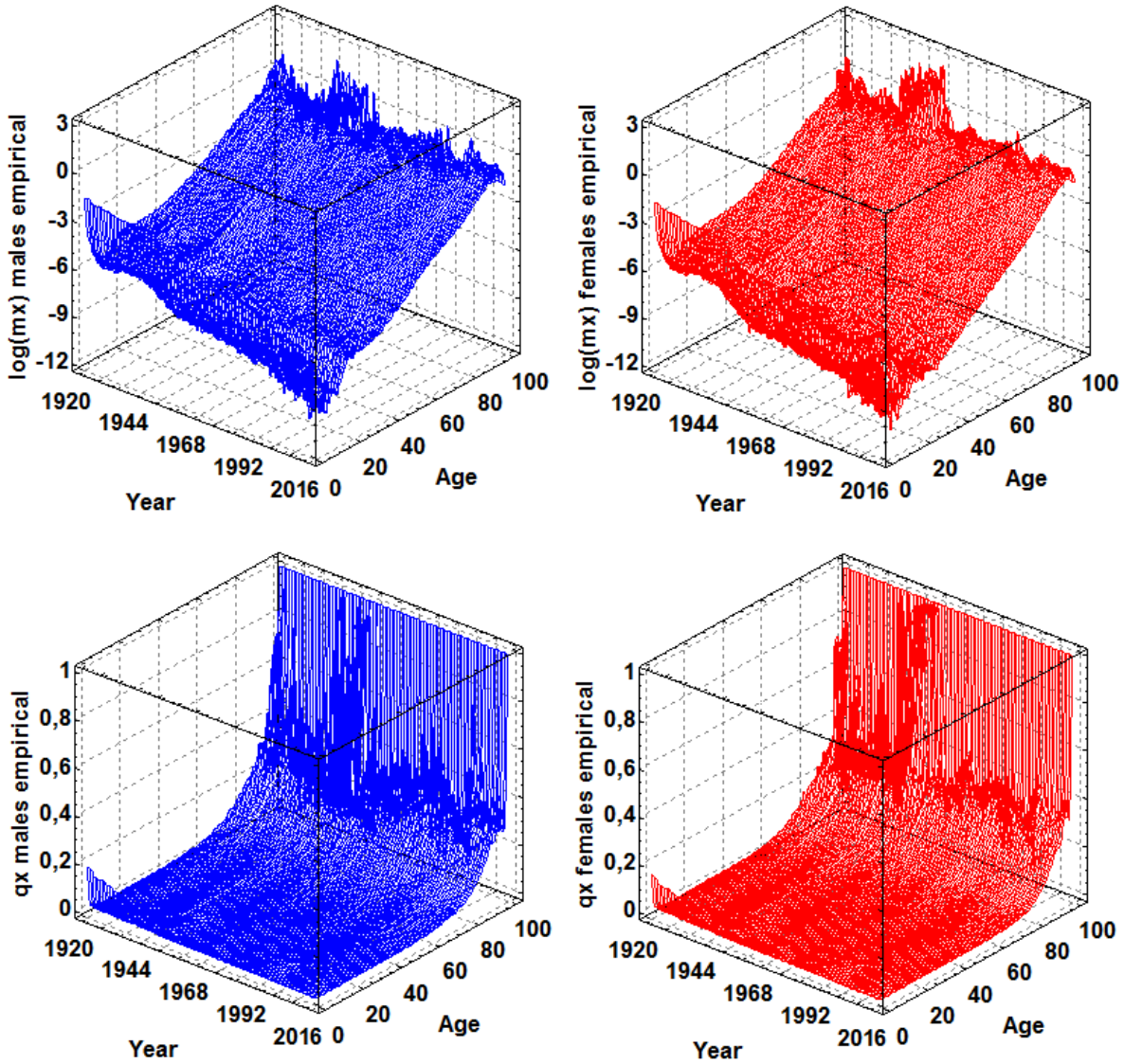


Figure 1 Logs of age-and-sex specific mortality rates for males and females in the Czech Republic (top charts) and calculated probabilities of death (bottom charts). Data source: CZSO, authors' illustration.

Overview of ARIMA models used on 2 × 101 time series is for males shown in the Table. 1, for females in Table 2. These models were evaluated by residual diagnostic tests (autocorrelation, heteroskedasticity and normality; see e.g. paper by Jarque and Bera [11]) and they are correct in all cases at 5% statistical significance level. It is clear from this overview that the most common form of models is ARIMA(0,1,1) with a constant. (This situation occurred in 77 cases (out of 101) in male population and in 61 cases (out of 101) in female population). This form was used as a compromise and a script was programmed to fit the 2 × 101 time series once again in the RStudio software. It was done this time without diagnostic tests, which could not be at 5% significance level statistically significant in some cases. Therefore, some parameters might have be deflected.

Age	Model	Age	Model	Age	Model	Age	Model	Age	Model	Age	Model	Age	Model	Age	Model
0	(0,2,1) c	13	(0,1,1) c	26	(0,1,1) c	39	(0,1,1) c	52	(0,1,1) c	65	(0,1,1) c	78	(0,1,1) c	91	(0,1,2) c
1	(0,1,1) c	14	(0,1,1) c	27	(0,1,1) c	40	(2,2,1)	53	(0,1,1) c	66	(0,1,1) c	79	(1,1,1) c	92	(2,1,0) c
2	(2,2,1)	15	(0,1,1) c	28	(0,1,1) c	41	(0,1,1) c	54	(0,1,1) c	67	(0,1,1) c	80	(0,1,1) c	93	(0,1,1) c
3	(0,1,1) c	16	(0,1,1) c	29	(0,1,1) c	42	(0,1,1) c	55	(0,1,1) c	68	(0,1,1) c	81	(0,1,1) c	94	(0,1,2) c
4	(0,1,1) c	17	(0,1,1) c	30	(0,1,1) c	43	(0,1,1) c	56	(2,1,0) c	69	(0,1,1) c	82	(0,1,1) c	95	(1,1,1) c
5	(0,1,1) c	18	(0,1,1) c	31	(2,2,1)	44	(0,1,1) c	57	(0,1,1) c	70	(0,1,1) c	83	(1,1,1) c	96	(0,1,2) c
6	(0,1,1) c	19	(0,1,1) c	32	(0,1,1) c	45	(0,1,1) c	58	(0,1,1) c	71	(0,1,1) c	84	(0,1,1) c	97	(0,1,1) c
7	(0,1,1) c	20	(0,1,2) c	33	(0,1,1) c	46	(0,1,1) c	59	(0,1,1) c	72	(0,1,1) c	85	(0,1,1) c	98	(0,1,1) c
8	(0,1,1) c	21	(0,1,1) c	34	(0,1,1) c	47	(2,1,0) c	60	(0,1,1) c	73	(1,1,0) c	86	(0,1,1) c	99	(0,1,2) c
9	(0,1,1) c	22	(0,1,1) c	35	(0,1,1) c	48	(1,1,0) c	61	(0,1,1) c	74	(1,1,0) c	87	(0,1,1) c	100+	(0,1,1) c
10	(0,1,1) c	23	(0,1,1) c	36	(0,1,1) c	49	(0,1,1) c	62	(0,1,1) c	75	(2,1,0) c	88	(0,1,1) c		
11	(0,1,1) c	24	(0,1,1) c	37	(0,1,1) c	50	(2,1,0) c	63	(0,1,1) c	76	(0,1,1) c	89	(0,1,2) c		
12	(0,1,1) c	25	(2,1,0) c	38	(0,1,1) c	51	(1,1,0) c	64	(0,1,1) c	77	(0,1,1) c	90	(2,1,0) c		

Table 1 ARIMA (*p,d,q*) models with or without constant for male's logarithms of age-specific mortality rates at the exact age 0–100+ in the Czech Republic. Source: authors' illustration.

Age	Model	Age	Model	Age	Model	Age	Model	Age	Model	Age	Model	Age	Model	Age	Model
0	(0,2,1) c	13	(0,1,1) c	26	(0,1,1) c	39	(0,1,1) c	52	(0,1,1) c	65	(1,1,0) c	78	(2,1,2) c	91	(1,1,1) c
1	(0,1,1) c	14	(1,1,1) c	27	(0,1,1) c	40	(2,1,0) c	53	(0,1,1) c	66	(1,1,0) c	79	(0,1,1) c	92	(2,1,0) c
2	(1,1,1) c	15	(0,1,1) c	28	(0,1,1) c	41	(1,1,0) c	54	(0,1,1) c	67	(0,1,1) c	80	(0,1,1) c	93	(0,1,1) c
3	(2,1,0) c	16	(2,1,0) c	29	(0,1,1) c	42	(0,1,1) c	55	(0,1,1) c	68	(0,1,1) c	81	(0,1,1) c	94	(2,1,1) c
4	(2,1,0) c	17	(0,1,1) c	30	(0,1,1) c	43	(0,1,1) c	56	(0,1,1) c	69	(0,1,1) c	82	(0,1,1) c	95	(1,0,0) c
5	(0,1,1) c	18	(0,1,2) c	31	(0,1,1) c	44	(0,1,1) c	57	(0,1,1) c	70	(0,1,1) c	83	(0,1,1) c	96	(1,0,0) c
6	(1,1,1) c	19	(0,1,1) c	32	(0,1,1) c	45	(0,1,1) c	58	(2,1,0) c	71	(1,1,0) c	84	(0,1,1) c	97	(0,0,2) c
7	(0,1,1) c	20	(0,1,1) c	33	(0,1,1) c	46	(0,1,1) c	59	(1,1,1) c	72	(2,1,0) c	85	(0,1,1) c	98	(0,0,2) c
8	(2,1,0) c	21	(0,1,1) c	34	(0,1,1) c	47	(0,1,1) c	60	(0,1,1) c	73	(1,1,0) c	86	(0,1,1) c	99	(1,1,0) c
9	(2,1,0) c	22	(2,1,0) c	35	(2,1,0) c	48	(0,1,1) c	61	(0,1,2) c	74	(2,1,0) c	87	(0,1,1) c	100	(0,1,1) c
10	(0,1,1) c	23	(0,1,1) c	36	(0,1,1) c	49	(1,1,1) c	62	(1,1,0) c	75	(2,1,0) c	88	(0,1,1) c		
11	(0,1,1) c	24	(1,1,0) c	37	(0,1,1) c	50	(0,1,1) c	63	(1,1,0) c	76	(0,1,1) c	89	(1,1,1) c		
12	(1,1,1) c	25	(0,1,1) c	38	(0,1,1) c	51	(0,1,1) c	64	(0,1,1) c	77	(0,1,1) c	90	(2,1,0) c		

Table 2 ARIMA (*p,d,q*) models with or without constant for female's logarithms of age-specific mortality rates at the exact age 0–100+ in the Czech Republic. Source: authors' illustration.

Optimized (user) models and compromised ARIMA (0,1,1) c (automatic) models were subsequently used to predict the indicator up to the year 2050. These mortality rates were recalculated using classical life table algorithm into the probability of death. Results are shown in the Figure 2, where at the top are the results of prediction by user models, in the middle part by automatic (ARIMA (0,1,1) c) models. For mutual comparison, the differences were calculated as

$$diff\left(q_{x,t}\right) = q_{x,t}^{user} - q_{x,t}^{automatic} \tag{5}$$

and these results are subsequently shown in the Figure 2 on the bottom. As can be seen from the results, there are almost no differences in prediction of the probability of death in the age groups of 0–95 years. Only in the highest age groups over 95 years there are deviations, which are probably caused due to variability in empirical data. If we would program a script that will smooth the input data using one of the existing levelling models, the ARIMA model (0,1,1) c would probably work better and the detected deviations would not occur. Because the variability in the highest age groups in input data is lower in the case of male population, the resulting deviations are therefore significantly smaller.

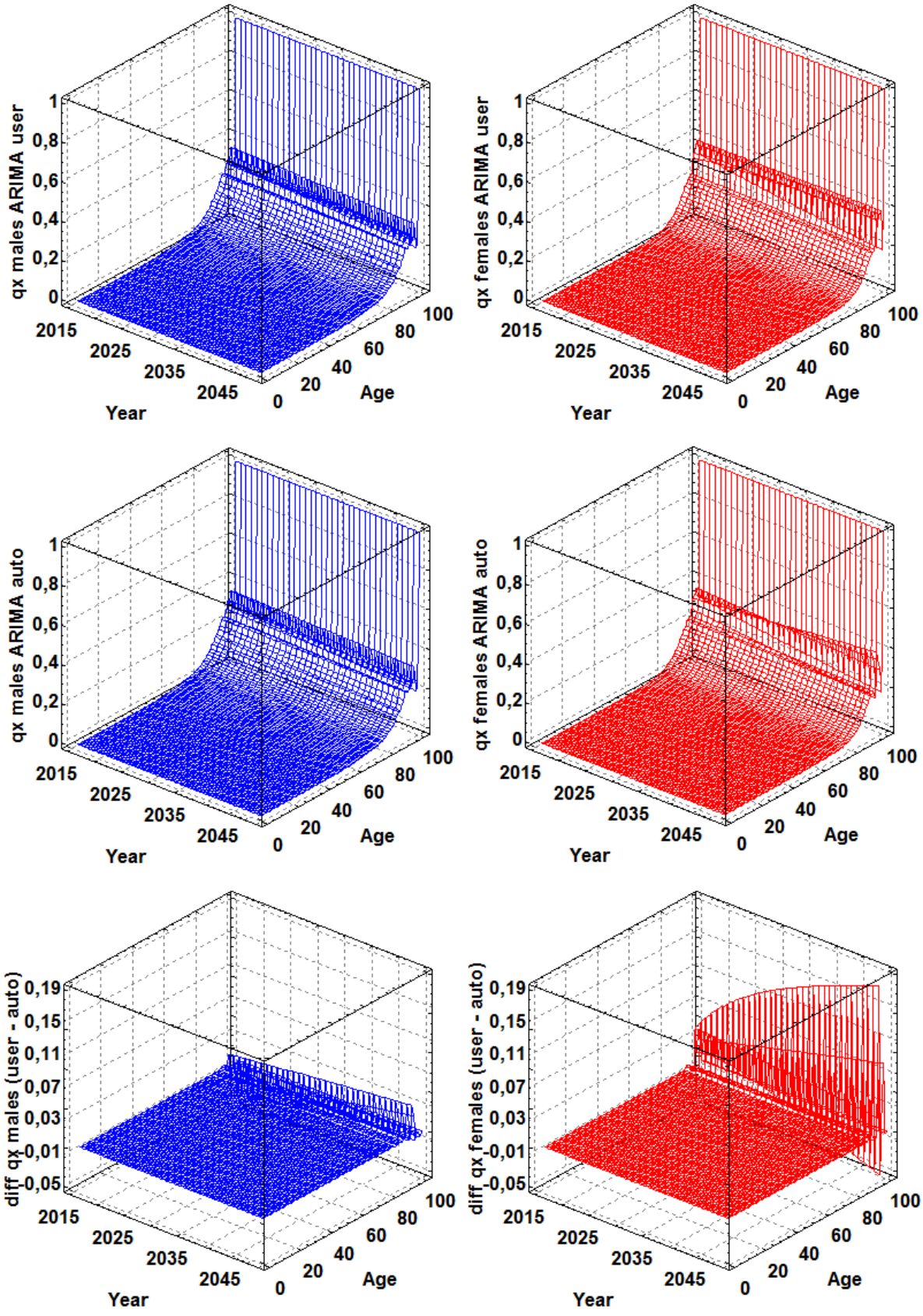


Figure 2 Forecasted age-and-sex-specific probabilities of deaths for male and female population in the Czech Republic up to the year 2050 using user and automatic approach. Source: authors' illustration.

4 Conclusion

The aim of this paper was to perform two approaches to modelling age-and-sex specific mortality rates for male and female population in the Czech Republic. We estimated 2 × 101 optimal ARIMA models and these subjected to the diagnosis of residues (autocorrelation, heteroskedasticity and normality). Then we forecasted mortality rates up to the year 2050. In the second step, we found that the most frequently occurring form of the model is ARIMA (0,1,1) with constant. We programmed script that uses this model structure and applies it to all 2 × 101 time series again using RStudio software, this time without the evaluation by the diagnostic tests. Consequently, we calculated the forecasts up to the year 2050 as well. It was found that in the case of male population are the results almost comparable. Differences depending on used model are greater in the case of female population. It is mainly caused by great variability in the raw data at the highest ages. This problem could be resolved by smoothing of mortality data by some of the existing models (see e.g. study by Gogola [8] or Godunov [9]), but that would bring a lot of extra steps that would completely eliminate the effect of saving work and effort.

Acknowledgements

The paper was supported from the Internal Grant Agency of University of Economics Prague no. 35/2017 "Demographic models in R Software".

References

[1] Andel, R. (2014). Aging in the Czech Republic. *Gerontologist*, vol. 54, no. 6, pp. 893-900.
[2] Antonion, K., Bardoutsos, A., and Ouburg, W. (2015). Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal*, vol. 5, no. 2, pp 245-281.
[3] Bell, W. R. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics*, vol. 13, no. 3, pp. 279-303.
[4] Box, G. E. P., Jenkins, G. (1970). *Time series analysis: Forecasting and Control*. San Francisco: Holden-Day.
[5] Fiala, T., Langhamrová, J. (2014). Increase of Labor Force of Older Age – Challenge for the Czech Republic in Next Decades. *Procedia Economics and Finance*, vol. 12, pp. 144-153.
[6] Finkelstein, M. S. (2005). Lifesaving explains mortality decline with time. *Mathematical Biosciences*, vol. 196, no. 2, pp. 187-197.
[7] Gardner Jr. E. S., McKenzie, E. (1985). Forecasting Trends in Time Series. *Management Science*, vol. 31, no. 10, pp. 1237-1246.
[8] Gogola, J. (2015). Modelling mortality rate of the very old. *In: 9th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Cracow: Cracow University of Economics, pp. 58-67.
[9] Godunov, I. (2015). *Comparison of official life tables construction in selected countries [Bachelor's thesis]*. Charles University in Prague, Faculty of Science, Department of Demography and Geodemography.
[10] Gompertz, B. (1825) On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London*, vol. 115, pp. 513-585.
[11] Jarque, C. M., Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, vol. 6, no. 3, pp. 255-259.
[12] Jindrová, P., Slavíček, O. (2014). Life Expectancy Development and Prediction for Selected European Countries *In: 6th International Scientific Conference Managing and Modelling of Financial Risks*, Ostrava: VŠB-TU Ostrava, pp. 303-312.
[13] Kirkwood, T. B. L. (2015). Deciphering death: a commentary on Gompertz (1825) 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'. *Philosophical Transactions of the Royal Society, B-Biological Sciences*, vol. 370, no. 1666, UNSP 20140379.
[14] Lee, R. D. Carter, L. (1992). Modeling and Forecasting the Time Series of U.S. Mortality. *Journal of the American Statistical Association*, vol. 87, pp. 659-671.
[15] Li, N., Lee, R. D. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, vol. 42, no. 3, pp. 575-594.
[16] Ševčíková, H., Li, N., Kantorová, V., Gerland, P., and Raftery, A. E. (2015). *Age-Specific Mortality and Fertility Rates for Probabilistic Population Projections*. Center for Statistics and the Social Sciences, University of Washington, Working Paper no. 150, March 17, 2015, pp. 1-31.
[17] Šimpach, O., Langhamrová, J. (2014). Stochastic Modelling of Age-specific Mortality Rates for Demographic Projections: Two Different Approaches. *In: 32nd International Conference on Mathematical Methods in Economics (MME)*, Olomouc: Palacký University in Olomouc, pp. 890-895.

Distance-based linguistic approximation methods: graphical analysis and numerical experiments

Tomáš Talášek[1], Jan Stoklasa[2]

Abstract. Linguistic approximation (LA) as a tool for converting the outputs of mathematical models into linguistic terms or expressions is a crucial tool in linguistic fuzzy modelling. The success of the models depends significantly on the ability of the users of these models to understand well enough the outputs provided by the models. Linguistic approximation offers a natural language for conveying information. On the other hand it is still an approximation of the original results and as such, there is information distortion taking place.
In this paper we study several distance-based linguistic approximation methods and analyse their performance in terms of LA for Mamdani-type outputs of mathematical models using a numerical experiment. We provide graphical summaries of the performance of these distance measures in LA as well as the frequencies of choosing specific linguistic labels considered to be the values of an extended linguistic scale. We discuss the differences in the focus of these methods and its implications for their usability. The paper strives to increase understanding of the LA methods and to contribute to the creation of a LA road map for practical use.

Keywords: Linguistic approximation, numerical experiment, distance-based methods, graphical analysis, Mamdani, fuzzy.

JEL classification: D81, C44
AMS classification: 90B50, 91B06

1 Introduction

Mathematical models for decision support and mathematical models representing expert knowledge frequently provide outputs in the form of fuzzy numbers or intervals (see e.g. [3, 7, 11, 15, 22]). In such cases when uncertainty is present, it might be convenient to provide the decision makers also with linguistic summaries of these results. Linguistic fuzzy modelling applying appropriate linguistic approximation (see e.g. [2, 6, 12, 25, 27] for some linguistic approximation techniques examples) can thus help to enhance the understandability of the outputs. Since the process of linguistic approximation can distort the information that is being approximated - note that the most fitting linguistic label from a usually small set of available well understood labels is assigned - a thorough investigation of the process of LA is needed. The research in this area has already started from the theoretical [4, 26] and behavioral [13, 16, 24] perspective and also from the perspective of the performance of LA methods in various contexts [18, 23, 19, 20]. Also alternative uses of LA were studied recently [14, 17, 21]. Most of the recent studies of LA methods and their applicability focus on the approximation of rather simple objects - i.e. fuzzy numbers (and frequently of the triangular or rectangular type). More general types of outputs, such as the outputs of Mamdani fuzzy inference [8] remain unaddressed. This paper strives to suggest the first step toward the analysis of the performance of LA methods in connection with Mamdani-type outputs.

2 Preliminaries

Let U be a nonempty set (the universe of discourse). A *fuzzy set* A on U is defined by the mapping $A : U \to [0, 1]$. A family of all fuzzy sets on U is denoted by $\mathcal{F}(U)$. For each $x \in U$ the value $A(x)$ is called a *membership degree* of the element x in the fuzzy set A and $A(.)$ is called a *membership function* of the fuzzy set A. $\text{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes a *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an α-*cut* of A for any $\alpha \in [0, 1]$, $\text{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A. Let A and B be fuzzy sets on the same universe U. We say that A is a *fuzzy subset* of B ($A \subseteq B$), if $A(x) \leq B(x)$ for all $x \in U$.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1) $\text{Ker}(A) \neq \emptyset$ (A is *normal*); (2) A_α are closed intervals for all $\alpha \in (0, 1]$ (this implies A is *unimodal*); (3) $\text{Supp}(A)$

[1]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic, tomas.talasek@upol.cz.

[2]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic and Lappeenranta University of Technology, School of Business and Management, Skinnarilankatu 34, Lappeenranta, Finland, jan.stoklasa@upol.cz.

is bounded. A family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$. A fuzzy number A is said to be defined on [a,b], if Supp(A) is a subset of the interval $[a,b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1,a_4] =$ Cl(Supp(A)) and $[a_2,a_3] =$ Ker(A), where Cl(Supp(A)) denotes a closure of Supp(A). Each fuzzy number A is determined by $A = \left\{[\underline{a}(\alpha), \overline{a}(\alpha)]\right\}_{\alpha\in[0,1]}$, where $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ is the lower and upper bound of the α-cut of fuzzy number A respectively, $\forall\alpha \in (0,1]$, and the closure of the support of A Cl(Supp(A)) $= [\underline{a}(0), \overline{a}(0)]$. The *cardinality* of fuzzy number A on $[a,b]$ is a real number Card(A) defined as follows: Card$(A) = \int_a^b A(x)dx$. A *union* of two fuzzy sets A and B on U is a fuzzy set $(A \cup B)$ on U defined as folllows: $(A \cup B)(x) = \min\{A(x), B(x)\}$ and a *Łukasiewicz union* of two fuzzy sets A and B on U is a fuzzy set $(A \cup_L B)$ on U defined as follows: $(A \cup_L B)(x) = \min\{1, A(x) + B(x)\}$, $\forall x \in U$. Let $A_1, \ldots, A_n$ be a fuzzy sets on $U_1, \ldots, U_n$ respectively. The *Cartesian product* of $A_1, \ldots, A_n$ is a fuzzy set $(A_1 \times \cdots \times A_n)$ on $U_1 \times \cdots \times U_n$ with membership function $(A_1 \times \cdots \times A_n)(x_1, \ldots, x_n) = \min\{A_1(x_1), \ldots, A_n(x_n)\}$, $\forall x_i \in U_i$. A fuzzy set R on $U_1 \times \cdots \times U_n$ is called an *n*-ary fuzzy relation. Let R be a fuzzy relation on $U \times V$ and S be a fuzzy relation on $V \times W$. The composition $(R \circ S)$ is a fuzzy set on $U \times W$ a with membership function $(R \circ S)(x,z) = \sup_{y\in V} \min\{R(x,y), S(y,z)\}$, $\forall x \in U, z \in W$.

The fuzzy number A is called *linear* if its membership function is linear on $[a_1,a_2]$ and $[a_3,a_4]$ and $a_1 \neq a_4$; for such fuzzy numbers we will use a simplified notation $A = (a_1,a_2,a_3,a_4)$. A linear fuzzy number A is said to be *trapezoidal* if $a_2 \neq a_3$ and *triangular* if $a_2 = a_3$. We will denote triangular fuzzy numbers by ordered triplet $A = (a_1,a_2,a_4)$. Triangular fuzzy number $A = (a_1,a_2,a_4)$ is called *symmetric triangular fuzzy number* if $a_2 - a_1 = a_4 - a_2$. More details on fuzzy numbers and computations with them can be found for example in [5].

A *fuzzy scale* on $[a,b]$ is defined as a set of fuzzy numbers $T_1, T_2, \ldots, T_s$ on [a,b], that form a Ruspini fuzzy partition (see [10]) of the interval $[a,b]$, i.e. for all $x \in [a,b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering. A *linguistic variable* ([27]) is defined as a quintuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$, where $\mathcal{V}$ is a name of the variable, $\mathcal{T}(\mathcal{V})$ is a set of its linguistic values (terms), X is an universe on which the meanings of the linguistic values are defined, G is an syntactic rule for generating the values of $\mathcal{V}$ and M is a semantic rule which to every linguistic value $\mathcal{A} \in \mathcal{T}(\mathcal{V})$ assigns its meaning $A = M(\mathcal{A})$ which is usually a fuzzy number on X. Linguistic variable $(\mathcal{V}, \mathcal{T}(\mathcal{V}), X, G, M)$ is called a *linguistic scale* on $[a,b]$ if $X = [a,b]$, $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ and $T_i = M(\mathcal{T}_i)$, $i = 1, \ldots, n$ form a fuzzy scale on $[a,b]$. Fuzzy scale is called *uniform* when Card(Supp(T_i)) $= 2 \cdot (b-a)/(n-1)$ for all $i = 2, ..., n-1$, Card(Supp(T_i)) $= (b-a)/(n-1)$ for $i = 1$ and $i = n$, T_i form a Ruspini fuzzy partition of U, and $T_2, \ldots, T_{n-1}$ are symmetrical triangular fuzzy numbers. Linguistic terms $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ of linguistic scale $\mathcal{T}(\mathcal{V})$ are called *elementary (level 1) terms* of linguistic scale. Linguistic scale using additional linguistic terms $\mathcal{T}_i$ to $\mathcal{T}_j$ where $i = 1, \ldots, n-1$, $j = 2, \ldots, n$ and $i < j$ (called *derived linguistic terms*) is called *extended linguistic scale*; $M(\mathcal{T}_i$ to $\mathcal{T}_j) = T_i \cup_L T_{i+1} \cup_L \cdots \cup_L T_j$. The extended linguistic scale thus contains linguistic values of different levels of uncertainty – from the possibly least uncertain elementary terms $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ to the most uncertain linguistic term $\mathcal{T}_1$ to $\mathcal{T}_n$ (uncertainty can be assessed by the cardinality of the meanings of these linguistic terms). Derived linguistic terms $\mathcal{T}_i$ to $\mathcal{T}_j$ are called *level* $j - i + 1$ *terms* and can be also denoted by $\mathcal{T}_{ij}$.

Let Out be a fuzzy number on $[a,b]$ and $(\mathcal{V}, \mathcal{T}(\mathcal{V}), [a,b], G, M)$ be a linguistic variable such that $\mathcal{T}(\mathcal{V}) = \{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$ and $T_i = M(\mathcal{T}_i)$, $i = 1, \ldots, s$, are fuzzy numbers on $[a,b]$. The *linguistic approximation* of fuzzy number Out is a process of searching for a suitable linguistic term $\mathcal{T}_{Out}$ from $\mathcal{T}(\mathcal{V})$ which describes the meaning of the fuzzy number Out the best. One of the most popular approaches to finding the linguistic term $\mathcal{T}_{Out}$ is using the "best-fit" approach:

$$T_{Out} = \arg \min_{i\in\{1,\ldots,s\}} d(T_i, Out), \tag{1}$$

where $d(A,B)$ is a distance measure[3] of two fuzzy numbers A and B.

3 Definition of a problem

As was outlined in the introduction, our investigation aims on the Mamdani-type outputs and their linguistic approximation. More specifically we aim on the linguistic approximation of the outputs of Mamdani fuzzy inference. Let us consider m linguistic scales $(\mathcal{V}_j, \mathcal{T}(\mathcal{V}_j), X_j, G_j, M_j)$, $j = 1, \ldots, m$, representing the inputs of the fuzzy inference system and an output linguistic scale $(\mathcal{W}, \mathcal{T}(\mathcal{W}), Y, G_{\mathcal{W}}, M_{\mathcal{W}})$. Let us also consider a collection of n rules representing the relationships between the input and output variables in the form:

$$\text{If } \mathcal{V}_1 \text{ is } \mathcal{A}_{i1} \text{ and } \ldots \text{ and } \mathcal{V}_m \text{ is } \mathcal{A}_{im}, \text{ then } \mathcal{W} \text{ is } \mathcal{B}_i,$$

where $\mathcal{A}_{j1} \in \mathcal{T}(\mathcal{V}_j)$ and $\mathcal{B}_i \in \mathcal{T}(\mathcal{W})$, $M_j(\mathcal{A}_{ij}) = A_{ij}$ and $M_{\mathcal{W}}(\mathcal{B}_i) = B_i$ for $i = 1, \ldots, n$, $j = 1 \ldots, m$. The output Out of Mamdani fuzzy inference computed for the input $(A'_1 \times \cdots \times A'_m)$, $A'_j \in \mathcal{F}(X_j)$, $j = 1, \ldots, m$,

[3] Alternatively a similarity measure of two fuzzy numbers can be used. In this case, the $\arg\min$ function in formula (1) is replaced by $\arg\max$.

using the fuzzy rule base R consisting of n fuzzy rules, $R = \bigcup_{i=1}^{n}(A_{i1} \times \cdots \times A_{im} \times B_i)$ is computed by (2).

$$Out = (A'_1 \times \cdots \times A'_m) \circ \bigcup_{i=1}^{n}(A_{i1} \times \cdots \times A_{im} \times B_i). \tag{2}$$

A sample output of this type is presented in Figure 1. Clearly Out does not need to be a convex fuzzy set any more. In this paper only normal Mamdani-type outputs are considered. The convexity of Out is not required. To find the linguistic approximation of Out, two distances of fuzzy sets that do not require convexity will be compared in terms of their performance in LA:

- *modified Bhattacharyya distance* [1]:

$$d_1(A, B) = \left[1 - \int_a^b (A^*(x) \cdot B^*(x))^{1/2} dx\right]^{1/2}, \tag{3}$$

  where $A^*(x) = A(x)/\text{Card}(A)$ and $B^*(x) = B(x)/\text{Card}(B)$,
- *Fuzzy distance* [9]:

$$d_2(A, B) = \frac{\int_a^b \mid A(x) - B(x) \mid}{\text{Card}(A) + \text{Card}(B)}. \tag{4}$$

4 Numerical analysis and discussion of the results

The numerical analysis of the performance of d_1 and d_2 in linguistic approximation of Mamdani-type outputs will be carried out under the assumption of a 5-element uniform linguistic scale representing the meanings of the elementary terms of the output variable by triangular fuzzy numbers $\{B_1, \ldots, B_5\} = \{(0, 0, 0.25), (0, 0.25, 0.5), (0.25, 0.5, 0.75), (0.5, 0.75, 1), (0.75, 1, 1)\}$ as presented in Figure 1. This assumption results in no loss of generality, as the results are generalisable for any number of elementary terms larger than 4, as long as their meanings form a uniform Ruspini partition on the given universe. More specifically an extended scale defined using elementary linguistic terms $\{\mathcal{B}_1, \ldots, \mathcal{B}_5\}$ is assumed for the purposes of LA. We also restrict our investigation to the linguistic approximation of the outputs the type $Out = h_L \cdot B_{k-1} \cup B_k \cup h_R \cdot B_{k+1}$ for $h_L, h_R \in [0, 1]$ and $k = 3, \ldots, t-2$, where t is the number of elementary terms in $\mathcal{W}$.

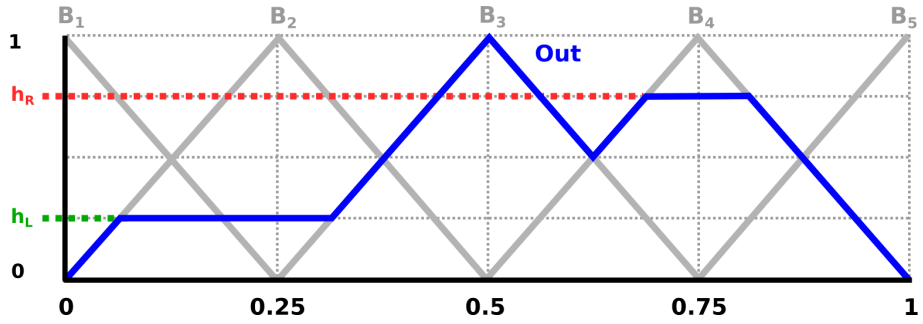


Figure 1: An example of a Mamdani-type output. This type of outputs of Mamdani fuzzy inference models is considered to be linguistically approximated in this paper. $B_1, \ldots, B_5$ are the meanings of the elementary linguistic values of the output linguistic variable used for the linguistic approximation.

For the purpose of systematic investigation of the behavior of LA under fuzzy distances d_1 and d_2, 501 uniformly distributed values of h_L and h_R from interval $[0, 1]$ were generated. This way, 251 001 Mamdani-type fuzzy sets $\{Out_1, \ldots, Out_{251001}\}$ were generated and linguistically approximated. Note, that each fuzzy set Out_i, $i = 1, \ldots, 251\,001$ can be unambiguously described by a 2-tuple (h_L, h_R). The results of the linguistic approximation using fuzzy distances d_1 and d_2 are depicted in Figures 2 and 3 respectively. The result of the linguistic approximation (i.e. the resulting element of the extended scale) for each approximated Mamdani-type output represented as a point with coordinates (h_L, h_R), is represented by a specific color.

It can by clearly seen from Figures 2 and 3 that although both fuzzy distances assign only four linguistic terms $\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}$ and $\mathcal{B}_{24}$, the results of LA are significantly different for each fuzzy distance. The Bhattacharyya distance favors linguistic terms the meanings of which are supersets to the linguistically approximated fuzzy set Out_i, $i = 1, \ldots, 251\,001$; only 1.33% of fuzzy sets are linguistically approximated by the elementary term $\mathcal{B}_3$ using the Bhattacharyya distance d_1 (in the case of fuzzy distance d_2 it is more than 15%). Moreover, when the

value of h_L or h_R is higher than 0.13, the term $\mathcal{B}_3$ is never assigned based on d_1. Higher uncertainty of Out thus implies that an at least level-2 term will be assigned as a linguistic approximation using the Battacharyya distance. Also linguistic terms $\mathcal{B}_{23}, \mathcal{B}_{34}$ are not used often – each of them is used in less than 9% of the cases. That is caused by the fact, that when both h_L and h_R are higher than 0.135 simultaneously, the only possible outcome of LA is $\mathcal{B}_{24}$ (i.e. a level-3 term). Actually, more than 80.7% of the approximated fuzzy sets are approximated by the term $\mathcal{B}_{24}$. This only confirms the findings obtained in [20] that Bhattacharyya distance tends to suggest such approximating labels that tend to be supersets of the approximated object meaning-wise.

On the contrary the fuzzy distance d_2 divided the space of the approximated fuzzy numbers Out_i, $i = 1, \ldots, 251\,001$ into four rectangular-like areas (see Figure 3) with respect to the result of the linguistic approximation. Note that all four possible outcomes of the linguistic approximation are suggested with similar frequencies. Linguistic terms $\mathcal{B}_{23}, \mathcal{B}_{34}$ are both assigned to almost 25% of the approximated fuzzy sets. Linguistic term $\mathcal{B}_{24}$ (the most uncertain term that is used) is assigned to 35% of approximated fuzzy sets. Note that for $h_L = h_R \approx 0.366$ we obtain a fuzzy set that can be linguistically approximated by all four linguistic terms $\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}$ and $\mathcal{B}_{24}$ (the fuzzy distance d_2 between this fuzzy set and the meaning of each of the four linguistic terms considered is the same). This could be a potential handicap of this fuzzy distance due to the fact, that just a small change of h_L and/or h_R around this point could result in a different linguistic label from the set $\{\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}, \mathcal{B}_{24}\}$. Only limited number of approximated fuzzy sets results into ambiguous cases, when the LA was unable to assign proper linguistic term (10 fuzzy sets in the case of d_1 and 22 in the case of d_2, see Table 1). These ambiguous cases are depicted in the Figures 2 and 3 by yellow color. Note, that in fact the borders between the pairs of differently colored areas are always constituted by ambiguous cases. Due to the chosen mesh just some of the ambiguous cases manifested themselves in the numerical analysis.

Thus if the meaning of the linguistic approximation is required to be a superset to the approximated fuzzy set, it is more reasonable to choose Bhattacharyya distance d_1 over d_2. The requirement of supersets is a "safe" approach to linguistic approximation, since the reduction of uncertainty is not significant (in fact the approximating fuzzy set is frequently more uncertain). This can, however, lead to the situations where most of the approximated fuzzy sets will be labeled by the same linguistic term. The fuzzy distance d_2 uses the available linguistic terms of the approximating extended linguistic scale more uniformly, which results in low-uncertain outputs being approximated by a low uncertain linguistic term (but under a larger risk that a part of the approximated fuzzy output will not be covered well by the selected linguistic approximation, i.e. the intersection of the meaning of the approximating term and the output will be nonempty).

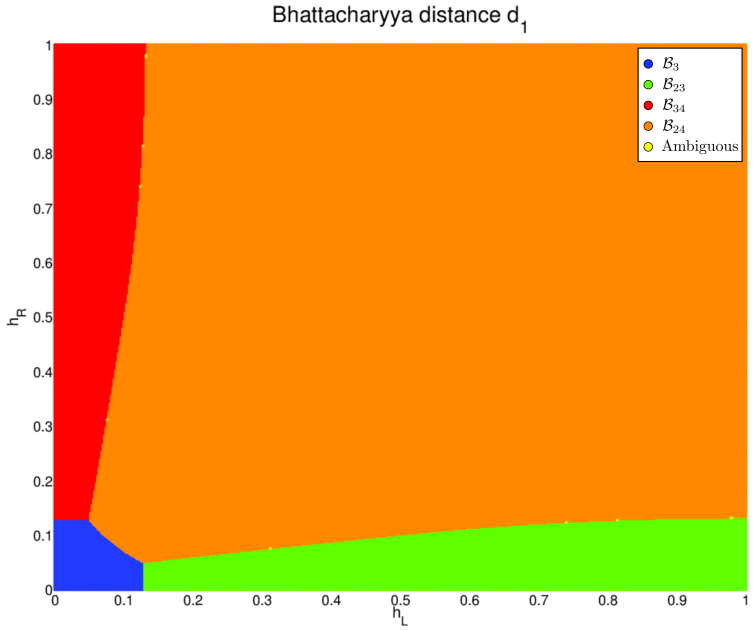


Figure 2: The results of the linguistic approximation of fuzzy sets $\{Out_1, \ldots, Out_{251001}\}$ using Bhattacharyya distance d_1. Approximated Mamdani-type fuzzy sets are represented by points with coordinates (h_L, h_R). The color of each point represents the linguistic approximation of the corresponding fuzzy set. Ambiguous cases (when more than one linguistic term is suggested based on d_1) are depicted using yellow color.

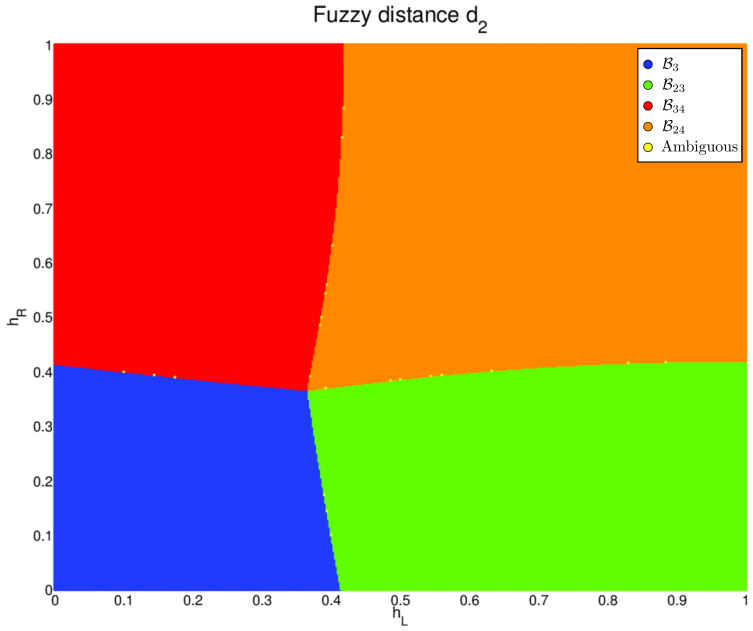


Figure 3: The results of the linguistic approximation of fuzzy sets $\{Out_1, \ldots, Out_{251001}\}$ using Fuzzy distance d_2. Approximated Mamdani-type fuzzy sets are represented by points with coordinates (h_L, h_R). The color of each point represents the linguistic approximation of the corresponding fuzzy set. Ambiguous cases (when more than one linguistic term is suggested based on d_2) are depicted using yellow color.

	$\mathcal{B}_3$	$\mathcal{B}_{23}$	$\mathcal{B}_{34}$	$\mathcal{B}_{24}$	**Ambiguous**
d_1	3 329	22 541	22 541	202 580	10
d_2	37 948	61 806	61 806	89 419	22

Table 1: Frequencies of assignment of linguistic terms $\{\mathcal{B}_3, \mathcal{B}_{23}, \mathcal{B}_{34}, \mathcal{B}_{24}\}$ and Ambiguous cases obtained for fuzzy sets $\{Out_1, \ldots, Out_{251001}\}$ in the linguistic approximation using d_1 and d_2. Unlisted linguistic terms (e.g. $\mathcal{B}_1, \mathcal{B}_{13}, \ldots$) were not assigned to any approximated Mamdani-type fuzzy set.

5 Conclusion

Mamdani-type fuzzy sets are most frequently obtained through Mamdani fuzzy inference systems. These systems are represented by a set of fuzzy IF-THEN rules which can be formulated linguistically. As such these systems present a useful tool for the descrtiption and investigation of economical phenomena and rather complex systems. This paper investigates the performance of linguistic approximation of Mamdani-type outputs [8] under two different fuzzy distance measures. The results of this numerical analysis clearly show, that the choice of the fuzzy distance measure can significantly affect the results of linguistic approximation. In the paper we have identified several important properties that could help the users of mathematical models choose the most suitable fuzzy distance measure. First, although it is not frequently studied in the literature, linguistic approximation of Mamdani-type fuzzy sets can be done, even within the "best-fit" LA context. Second, performance of the LA can be customized by an appropriate choice of the distance measure - d_1 is more prone to suggest a superset meaningwise, d_2 suggests LA closer in terms of cardinality. This paper is the initial step of exploring the results of LA of Mamdani-type outputs under different fuzzy distance/similarity measures.

Acknowledgements

Supported by the grant No. IGA_FF_2017_011 of the Internal Grant Agency of Palacký University Olomouc.

References

[1] Aherne, F., Thacker, N., and Rockett, P.: The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data. *Kybernetika* **32** (1998), 363–368.

[2] Bonissone, P. P.: A fuzzy sets based linguistic approach: theory and applications. In: *Proceedings of the 12th conference on Winter simulation*, 1980. 99–111.

[3] Collan, M., Stoklasa, J., and Talašová, J.: On Academic Faculty Evaluation Systems: More than just Simple Benchmarking. *International Journal of Process Management and Benchmarking* **4** (2014), 437–455.

[4] Degani, R., and Bortolan, G.: The problem of linguistic approximation in clinical decision making. *International Journal of Approximate Reasoning* **2** (1988), 143–162.

[5] Dubois, D., and Prade, H., eds.: *Fundamentals of Fuzzy Sets*. Kluwer Academic Publishers, Massachusetts, 2000.

[6] Eshragh, F., and Mamdani, E. H.: A general approach to linguistic approximation. *International Journal of Man-Machine Studies* **11** (1979), 501–519.

[7] Krejčí, J., and Stoklasa, J.: Fuzzified AHP in the evaluation of scientific monographs. *Central European Journal of Operations Research* **24** (2016), 353–370.

[8] Mamdani, E., and Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies* **7** (1975), 1 – 13.

[9] Pappis, C. P., and Karacapilidis, N. I.: A comparative assessment of measures of similarity of fuzzy values. *Fuzzy Sets and Systems* **56** (1993), 171 – 174.

[10] Ruspini, E. H.: A new approach to clustering. *Information and control* **15** (1969), 22–32.

[11] Stoklasa, J.: Multiphase linguistic fuzzy model for the Czech emergency medical rescue services. In: *Proceedings of the 28th International Conference on Mathematical Methods in Economics*, 2010. University of South Bohemia in České Budějovice, 590–595.

[12] Stoklasa, J.: *Linguistic models for decision support*. Lappeenranta University of Technology, Lappeenranta, 2014.

[13] Stoklasa, J., and Talášek, T.: Jazykově orientované modelování pro ekonomickou a manažerskou praxi některé otevené problémy. In: *Znalosti pro tržní praxi 2015*. 959–969.

[14] Stoklasa, J., and Talášek, T.: On the use of linguistic labels in AHP: calibration, consistency and related issues. In: *Proceedings of the 34th International Conference on Mathematical Methods in Economics*. Technical University of Liberec, Liberec, 785–790.

[15] Stoklasa, J., Talášek, T., Kubátová, J., and Seitlová, K.: Likert scales in group multiple-criteria evaluation. *Journal of Multiple-Valued Logic and Soft Computing* (2017), (in press).

[16] Stoklasa, J., Talášek, T., and Musilová, J.: Fuzzy approach - a new chapter in the methodology of psychology? *Human Affairs* **24** (2014), 189–203.

[17] Stoklasa, J., Talášek, T., and Stoklasová, J.: Semantic differential and linguistic approximation - identification of a possible common ground for research in social sciences. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016*. Societas Scientiarum Olomucensis II, Olomouc, 495–501.

[18] Talášek, T., and Stoklasa, J.: Jazykově orientované modelování pro ekonomickou a manažerskou praxi - vybrané metody pro jazykovou aproximaci. In: *Znalosti pro tržní praxi 2015*. 977–989.

[19] Talášek, T., and Stoklasa, J.: The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016*. Societas Scientiarum Olomucensis II, Olomouc, 539–546.

[20] Talášek, T., and Stoklasa, J.: A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation Under Different Linguistic Scales. *Journal of Multiple-Valued Logic and Soft Computing* (2017), (in press).

[21] Talášek, T., Stoklasa, J., Collan, M., and Luukka, P.: Ordering of fuzzy numbers through linguistic approximation based on Bonissone's two step method. In: *16th IEEE International Symposium on Computational Intelligence and Informatics*. 285–290.

[22] Talášek, T., Stoklasa, J., and Talašová, J.: Linguistic approximation using fuzzy 2-tuples in investment decision making. In: *Proceedings of the 33rd International Conference on Mathematical Methods in Economics*. 817–822.

[23] Talášek, T., Stoklasa, J., and Talašová, J.: The role of distance and similarity in Bonissone's linguistic approximation method a numerical study. In: *Proceedings of the 34th International Conference on Mathematical Methods in Economics*. Technical University of Liberec, Liberec, 845–850.

[24] Talašová, J., Stoklasa, J., and Holeček, P.: HR management through linguistic fuzzy rule bases - a versatile and safe tool? In: *Proceedings of the 32nd International Conference on Mathematical Methods in Economics*. 1027–1032.

[25] Wenstøp, F.: Quantitative analysis with linguistic values. *Fuzzy Sets and Systems* **4** (1980), 99–115.

[26] Yager, R. R.: On the retranslation process in Zadeh's paradigm of computing with words. *IEEE transactions on systems, man, and cybernetics. Part B: Cybernetics* (2004), 1184–1195.

[27] Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences* **8** (1975), 199–249.

Mathematical support for human resource management at universities

Jana Talašová [1], Jan Stoklasa[2], Pavel Holeček [3], Tomáš Talášek[4]

Abstract. Universities in the Czech Republic are required by law to continuously assess and ensure their quality in a verifiable manner. The quality of university staff, primarily of academic staff, is one of the most important aspects of the quality of a university. This paper describes the current state of development of the information system for academic staff evaluation (IS HAP) and for the evaluation of other university staff (IS HOP) with the focus on evaluation methods.
In the IS HAP the use of objectively verifiable information on the activities of academic staff members and their results in the teaching and research is stressed. Input data is acquired from verifiable sources and evaluated against standards determined for different work positions and evaluated areas. The aggregation of partial evaluations is performed by a fuzzy rule base. The calculated evaluations are represented verbally and using colours. The superiors then use these results to derive the overall narrative evaluations.
The IS HOP is based on the assessment of the main areas of activities of other staff by their superiors, linguistic scales are used in their evaluation and colour representation of outputs is used.

Keywords: Human resource management, university, evaluation models, fuzzy sets, information systems.

JEL classification: C44
AMS classification: 90C15

1 Introduction

The quality of human resources is a key factor influencing the quality of a university (see [5]). As far as academic staff is concerned, there is a large amount of information that can and should be used while managing them (data concerning the teaching activities of academic staff, data concerning their creative activity and also data on other activities important for the well-being and development of the university). The Structure of applied data varies from university to university (see e.g. [1, 2, 3, 4, 6, 8]). A great deal of the data is currently readily available in various information systems of universities, there are, however, also additional pieces of data worth gathering (as these can be further used by the university itself). It is reasonable to create a specialized information system for the purpose of managing of academic staff. Considering the amount of data required and used for this purpose, it is necessary to not only gather the data in one information system and make it available in full details for further use, but also to process it within the system and to provide the university management with aggregated information useful for the management of human resource. The mathematical model used for the evaluation should be understandable to the evaluators as well as to the people that are being evaluated and most of all it should provide easily and clearly interpretable outputs. The information system should also be a platform for storing the final evaluation report in which the superior reflects both the information provided by the evaluation model as well as his/her own knowledge of the academic staff member and the relevant context. These are the main principles on which the *Information system for academic staff evaluation* (IS HAP) is based (see [7, 10, 12, 13]). This system was created in 2012 at Palacký University Olomouc and it has been continuously further developed since then. The quality of other (non-academic) university staff is also important to ensure proper functioning of the university. Considering the variety of professions of other academic staff, it is more appropriate to use a simpler model for the evaluation. Thus a unified questionnaire can be used by the superiors for the assessment of all non-academic staff members under their supervision concerning the fulfilment of their duties. Again, the clarity of outputs is important for efficient management – mathematical tools are thus again required for the processing and evaluation of data. The

[1]Palacký University Olomouc, Strategic Development Division, Křížkovského 8, Olomouc, Czech Republic, jana.talasova@upol.cz.

[2]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic and Lappeenranta University of Technology, School of Business and Management, Skinnarilankatu 34, Lappeenranta, Finland, jan.stoklasa@upol.cz.

[3]Palacký University Olomouc, Computer Center, Biskupské nám. 1, Olomouc, Czech Republic, pavel.holecek@upol.cz.

[4]Palacký University Olomouc, Department of Applied Economics, Křížkovského 12, Olomouc, Czech Republic, tomas.talasek@upol.cz.

information system for the evaluation of other university staff (IS HOP) created at Palacký University in 2015 offers a solution to the problem of evaluation of other university staff.

2 Evaluation of academic staff (IS HAP)

2.1 Input data for the evaluation model

Data considering the annual performance of academic staff is inserted in the IS HAP input form. The form is divided into three main areas: teaching activities, creative activities (incl. R&D) and academic offices and managerial duties. The performance evaluation is based on the first two areas; the third area is used to assess the total time-/work-load of the academic staff member. The area of teaching activities (TA) is further divided into three subareas: a) direct teaching, b) supervision of students, c) activities associated with the development of fields of study. The area of creative activities (CA) is divided into the following subareas: a) outputs resulting in RIV points (according to the Czech methodology for the evaluation of scientific outputs), b) other research outputs, c) administrative activities associated with creative activities, d) outputs resulting in RUV points (according to the methodology for the evaluation of creative work outcomes of art universities and faculties). The objectivity of input data is stressed significantly – mainly basic activities for which information can be obtained from reliable and verifiable sources are reported. Additional information provided by the academic staff members themselves is expected to be checked by their superior.

2.2 The evaluation model and its outputs

The selection of activities to be reflected in the evaluation process is a part of the creation of the evaluation model. The selected activities are subsequently assigned points. In the area of teaching activities, the points reflect time consumption and professional requirements. Time-consumption is, however, not a factor in the area of creative activities; excellence of the output is stressed. The basis (and universal benchmark) for the point-evaluation of activities in CA is the amount of points obtained for a paper published in an impact-factor journal, as computed by the following formula

$$b = 10 + 295 \cdot \frac{1 - N}{1 + \frac{N}{0.057}}, \quad N = 1 - \frac{AP}{100},$$

where b is the amount of points assigned for the paper and AP is the average percentile of the journal with respect to the impact factor across all the clusters in which the journal is indexed in the Web of Science. The scores of other activities in this area were determined by comparison with b, in a similar way as it was done in the methodology for the evaluation of outputs of Research, Development and Innovations activities valid in the Czech Republic in the recent years. The evaluations in RUV are determined on the same scale (point-wise).

Evaluation scores of an academic staff member in TA and CA areas are computed as sums of points assigned to the activities the staff member has reported in the given area. It is also possible to set the maximum frequency with which certain activities will be reflected in the overall evaluation score in the given area. This is intended to prevent situations, where a large number of low-point activities would overweight crucial activities and outputs. Its purpose is also to weaken the possible influence of difficult-to-verify data provided by the academic staff members on the overall score in the given area.

The overall scores for TA and CA are standardized with respect to the academic position of the staff member by the following formulas:

$$ta = \frac{b_{TA}}{s_{TA}}, \quad ca = \frac{b_{CA}}{s_{CA}},$$

where ta is the standardized evaluation in TA area and ca is the standardized evaluation in the CA area for each academic staff member on the given position, b_{TA} and b_{CA} are the overall evaluations of the given staff member (point gains) in TA and CA area respectively, and s_{TA} and s_{CA} are the standard scores for the given area and position.

The standard scores are set up during the implementation of IS HAP for a given university/faculty based in the analysis of the data stored in IS HAP and also reflecting the goals of the management of the university/faculty. Statistical tools intended to facilitate the set-up process of standard scores are available directly in IS HAP.

It is necessary to reflect the different character of the scales used for the assessment of performance in TA and in CA (stemming from the different grounds for the determination of scores of individual activities and from different possibilities of getting higher multiples of standard scores in each area of activities) when calculating the overall aggregated evaluation of a given academic staff member. Instead of using some of the usual aggregation methods which assume the same character of the evaluation scales (i.e. weighted average), IS HAP uses the tools of linguistic fuzzy modelling (see [9]) for the aggregation of partial evaluations – linguistic fuzzy scales, fuzzy rule bases and fuzzy inference. Linguistic fuzzy scales currently used in the IS HAP for the interpretation of evaluations

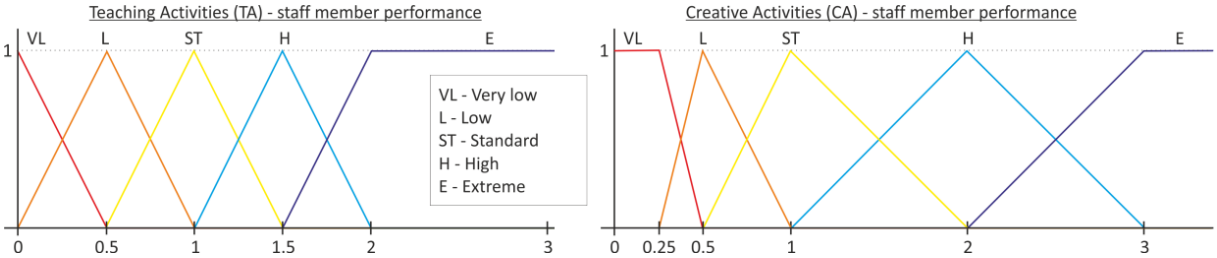


Figure 1: Linguistic fuzzy scale for the assessment of performance of a given academic staff member in TA (left) and CA (right) in IS HAP. Units of the x-axis are multiples of standard scores.

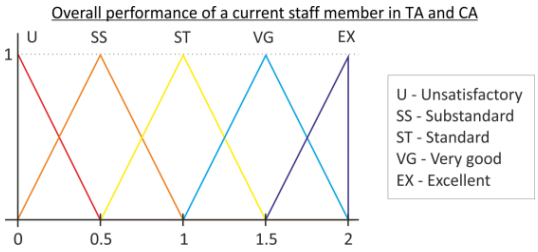


Figure 2: Linguistic fuzzy scale for the interpretation of the overall aggregated evaluation of a given staff member in TA and CA in IS HAP.

in TA and CA areas are depicted in Figure 1, the linguistic scale used for the interpretation of overall performance evaluation of a given academic staff member is depicted in Figure 2. The form of the fuzzy rule base used for the aggregation of partial evaluations in TA and CA areas into an overall performance evaluation depends on the type of faculty; Figure 3 presents a fuzzy rule base used for faculties with a balanced performance in teaching and creative activities, where a lower performance in one of the evaluated areas can, to some extent, be compensated by a higher performance in the other area of activities. When the Sugeno-Yasukawa fuzzy inference [11] is used, the aggregation function derived from the rule base assumes the following form:

$$eval(ta, ca) = \frac{\sum_{j=1}^{25} A_j(ta) \cdot B_j(ca) \cdot ev_j}{\sum_{j=1}^{25} A_j(ta) \cdot B_j(ca)} = \sum_{j=1}^{25} A_j(ta) \cdot B_j(ca) \cdot ev_j,$$

where $eval(ta, ca)$ represents the overall performance evaluation of a given academic staff member computed from his/her standardized partial evaluations ta and ca in TA and CA areas respectively, $A_j(.), B_j(.), j = 1, 2, \ldots, 25$, are all the combinations of pairs of membership functions of the elements of the two fuzzy scales depicted in Figure 1, and ev_j is the value from the kernel of a fuzzy set representing the overall evaluation (see Figure 2 for the 5-element linguistic scale used) corresponding with the j-th combination of fuzzy evaluations in TA and CA areas as specified by the fuzzy rule base depicted in Figure 3.

The output of the above described evaluation model is a real number. However, considering the imprecisions of the overall evaluations (stemming from the fact that all the possibly relevant activities can never be captured by the input form) and also for better interpretability of the outputs, we prefer linguistic and colour representation of the outputs of the evaluation model. A 5-element linguistic scale defined on the $[0, 2]$ universe of discourse of

Overall performance of an academic staff member in TA and CA		Creative Activities performance				
		Very low	Low	Standard	High	Extreme
Teaching Activities performance	Very low	Unsatisfactory	Unsatisfactory	Substandard	Standard	Very good
	Low	Unsatisfactory	Unsatisfactory	Substandard	Very good	Excellent
	Standard	Substandard	Substandard	Standard	Very good	Excellent
	High	Standard	Very good	Very good	Excellent	Excellent
	Extreme	Very good	Excellent	Excellent	Excellent	Excellent

Figure 3: Linguistic description of the fuzzy rule base for the aggregation of evaluations in TA and CA used in IS HAP.

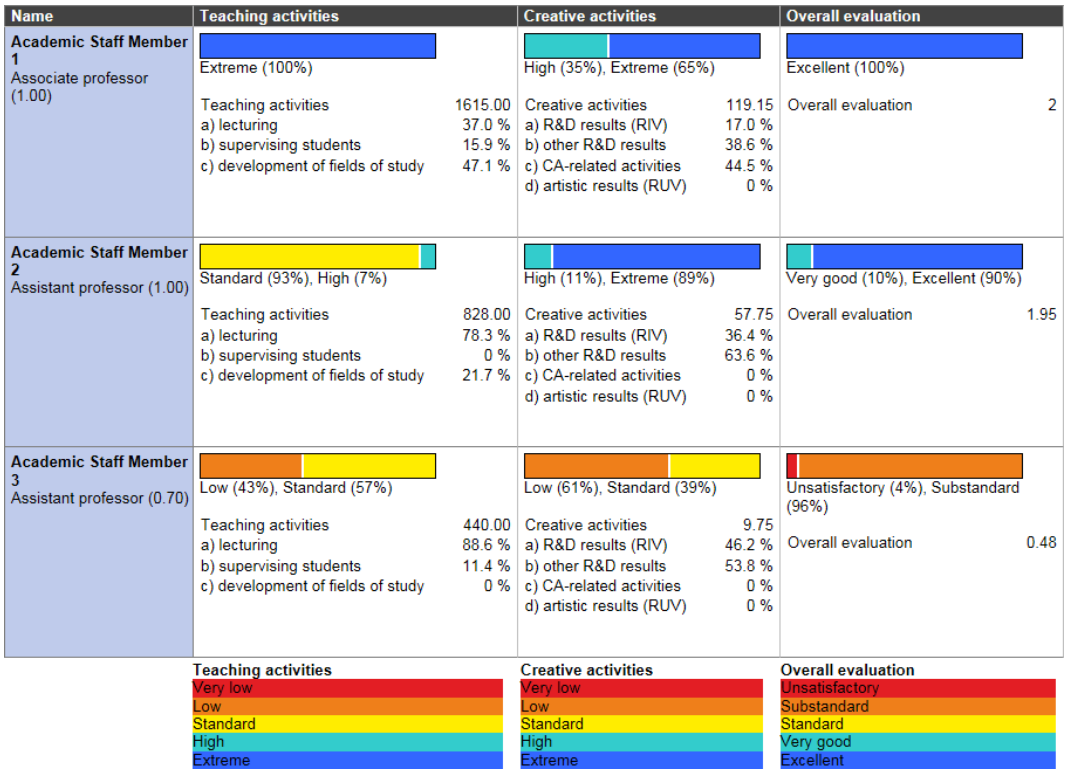


Figure 4: Overview of the evaluations of academic staff members.

the output variable (depicted in Figure 2) is therefore used. Using this linguistic scale, a real-number final output of the evaluation of e.g. 1.6 is thus interpreted as *80% very good* and *20% excellent*, or alternatively using a graphical representation by a rectangle which is 80% light-blue and 20% dark-blue. Partial evaluations in TA and CA areas are represented analogously. The graphical representation also allows for a comprehensive summary of the performance of staff of the whole unit (e.g. department) – see Figure 4.

2.3 Final evaluation of an academic staff member

The IS HAP is a tool for decision support; using this tool, the superiors of the evaluated academic staff members are provided with "hard" data concerning the activities of the academic staff member as well as with an aggregated information characterising his/her performance. The superiors are expected to add "soft" data concerning the given academic staff member relevant for the evaluation and reflect it in the final step of the evaluation process – the evaluation interview. The final narrative evaluation resulting from the evaluation interview is put by the superior into the respective field in IS HAP for storage.

3 Evaluation of other university staff (IS HOP)

3.1 Evaluation questionnaire and its outputs

Unlike academic staff, where the evaluation is based on a large amount of "hard" data and its sophisticated aggregation, other university staff is evaluated by a much simpler model (it is desirable to make the process simpler and the data requirements lower). Nevertheless, also the evaluation of other staff carried out regularly across a long period of time and stored in a specialized information system is an important tool for the management of university human resource. The IS HOP designed for this purpose therefore shares some features with IS HAP – its webpages look similarly and it uses the same organisational structure of the university and the same mechanisms for the identification of staff members. The main difference between the systems lies in the fact that in IS HOP the evaluation is carried out by the superiors of other university staff directly – by a selection of one of the pre-defined answers to the questions (items) of the evaluation questionnaire. The items included in the questionnaire are dependent on the specific requirements of the university/faculty/unit. The questionnaire can include e.g. just basic items concerning the quality of work, attitude to professional development, teamwork and approach to clients (a suggested solution of this type is presented in Figure 5). IS HOP, however, also supports the use of more complex evaluation questionnaires with different numbers of elements of the evaluation scales for each item. The same established way of presenting results as in IS HAP – i.e. graphical representation using colours – is adopted also in IS HOP. The

Figure 5: A screenshot from IS HOP – an example of evaluation questionnaire.

Name	Evaluations according to the individual criteria	
Employee 1 Economics Department Technical/office staff (1.00)	Professional knowledge and skills	B - very good
	Economical behavior	B - very good
	Team work, communication, encouragement of good relationship at the workplace	B - very good
	Attitude towards further professional development	C - standard
	Professional approach to clients	A - excellent
Employee 2 Economics Department Technical/office staff (1.00)	Professional knowledge and skills	B - very good
	Economical behavior	B - very good
	Team work, communication, encouragement of good relationship at the workplace	A - excellent
	Attitude towards further professional development	B - very good
	Professional approach to clients	A - excellent
Employee 3 Economics Department Technical/office staff (1.00)	Professional knowledge and skills	B - very good
	Economical behavior	C - standard
	Team work, communication, encouragement of good relationship at the workplace	E - unsatisfactory
	Attitude towards further professional development	C - standard
	Professional approach to clients	D - substandard

Figure 6: A screenshot from IS HOP – overview of non-academic staff members performance evaluations.

calculation of colours of elements of evaluation scales with different numbers of values represents the mathematical aspect of this model. The standard presentation of outputs in IS HOP is in the form of a linguistically labelled colour profile of each non-academic staff member (see Figure 6). The aggregation of the partial evaluations into an overall evaluation (by a suitable aggregation function) is, however, also possible if required.

4 Conclusion

The system for the evaluation of academic staff performance IS HAP has undergone a rapid development since its initial creation. This included, among other things, the connection to external sources of information and a development of the underlying mathematical evaluation model. Based on the experience with the implementation of IS HAP on different types of faculties of five Czech universities, modifications were made to the evaluation linguistic fuzzy scales for the performance in the evaluated areas and in the linguistic scale used for the interpretation of the overall evaluation. Several alternative rule bases were also developed to be used with different types of faculties and university units. The evaluation methodology of IS HAP has also been significantly improved, namely in terms of the integration of IS HAP into the internal system for quality assessment at universities. The system for the evaluation of other (non-academic) university staff IS HOP is in an earlier stage of development, its place in the internal system for quality assessment at universities is, however, also undeniable.

Acknowledgements

The research was supported by the grant IGA_FF_2017_011 of the Internal grant agency of Palacký University Olomouc.

References

[1] Faculty and Academic Staff Handbook of University of Wisconsin Oshkosh. `http://www.uwosh.edu/provost/resources/handbooks/facultyhandbook/2011%20Handbook%20FINAL.pdf`,

2011. [Online; accessed 23-November-2011].

[2] Faculty Handbook of A&M University. http://www.tamuk.edu/senate/_pdf_files/handbookaugust2011.pdf, 2011. [Online; accessed 23-November-2012].

[3] Guidelines for Evaluation of Faculty. http://facaffairs.med.wayne.edu/pdfs/2011_faculty_guidelines.pdf, 2011. [Online; accessed 23-November-2011].

[4] UTS Performance rating guidelines, Human Resources. http://www.hru.uts.edu.au/performance/reviewing/index.html, 2011. [Online; accessed 23-November-2011].

[5] Collan, M., Stoklasa, J., and Talašová, J.: On Academic Faculty Evaluation Systems: More than just Simple Benchmarking. *International Journal of Process Management and Benchmarking* **4** (2014), 437–455.

[6] Board of the Turku School of Economics. *Decision Regarding the Research Points Evaluation System*. Turku School of Economics, Turku, 2004.

[7] Holeček, P., Stoklasa, J., and Talašová, J.: Human resources management at universities - a fuzzy classification approach. *International Journal of Mathematics in Operational Research* **9** (2016), 502–519.

[8] Lappeenranta University of Technology. *Sisäinen ohje (Internal Instructions Regarding the Faculty Evaluation)*, LUT, Lappeenranta, 2011. [Dated: 24-March-2011].

[9] Stoklasa, J.: *Linguistic models for decision support*. Lappeenranta University of Technology, Lappeenranta, 2014.

[10] Stoklasa, J., Talašová, J., and Holeček, P.: Academic Staff Performance Evaluation - Variants of Models. *Acta Polytechnica Hungarica* **8** (2011), 91–111.

[11] Sugeno, M., and Yasukawa, T.: A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems* **1** (1993), 7–31.

[12] Talašová, J., and Stoklasa, J.: Fuzzy approach to academic staff performance evaluation. In: *Proceedings of the 28th International Conference on Mathematical Methods in Economics*. University of South Bohemia in České Budějovice, 621–626.

[13] Talašová, J., Stoklasa, J., and Holeček, P.: HR management through linguistic fuzzy rule bases - a versatile and safe tool? In: *Proceedings of the 32nd International Conference on Mathematical Methods in Economics*. Palacký University Olomouc, 1027–1032.

Transportation Problem Model Supplemented with Optimisation of Vehicle Deadheading and Single Depot Parking

Dušan Teichmann[1], Michal Dorda[2], Denisa Mocková[3]

Abstract. The transportation problem belongs to one of specialized problems of operations research. Its basic form is devoted to optimisation of transportation plans – how to supply customers from storehouses with homogeneous types of consignments. Optimisation criterions which are most often used are the total transportation costs or the total covered distance. In practice, we can find many possible applications of the transportation problem. For example, in rail transport the transportation problem can be applied in optimisation of capacity smoothing, in road transport in optimisation of transportation of empty containers between combined transport terminals and their customers and so on. To fulfil each transportation plan, we need some vehicles, the vehicles are assigned to a depot (or depots) where they are parked. The transportation plan usually consists of three types of trips. The first type trips are represented by the trips when the vehicles are loaded (the productive trips). The trips of the second and third type are non-productive (deadheading). The second type trips correspond to the trips of the empty vehicles going from the places of their unloading to the places where the vehicles are loaded again. The third type trips are the trips between the vehicle depots and the first loading places or the last places of unloading and the depots. However, the basic model of the transportation problem does not consider these second and third type trips.
An isolated solution for the individual types of the trips of such transportation problem is not suitable because decomposition does not assure optimality of such solution. The presented article presents a model in which all the mentioned trip types are mutually interconnected. That means a solution got by the mathematical model is optimal. The article is focused on an example with any number of the vehicles, all the vehicles have the same depot where they are parked and the capacity of all the vehicles is assumed to be 1. That means, for example, each vehicle can carry a container.

Keywords: Optimisation, mathematical programming, operational research, transportation problem, container system.

JEL Classification: C61
AMS Classification: 90C90

1 Introduction

The transportation problem is one of well-known problems of operations research. Its definition can be found in many publications devoted to mathematical programming – see for example [1], [2] or [6]. The transportation problem deals with decision how homogeneous consignments should be transported from storehouses to customers – that means we want to create a feasible transportation plan. Input data that is necessary for solving the transportation problem is given by capacities of the individual storehouses (how many consignments are found in each storehouse), demands of the individual customers (how many consignments should be delivered to each customer) and a matrix of unit transportation costs (elements of the matrix express how much it costs to transport the homogeneous consignment between the individual storehouses and the individual customers).

This is the basic formulation of the transportation problem. However, in practice it is often necessary to interconnect the created transportation plan with vehicle scheduling. The feasible schedule of the vehicles is usually planned in the following manner. The vehicle starts its schedule in a given place (usually a depot where the vehicle is parked), within its daily schedule the vehicle serves all the planned customers (that means it transports

[1] VSB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, Czech Republic, e-mail: dusan.teichmann@vsb.cz
[2] VSB – Technical University of Ostrava / Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, Czech Republic, e-mail: michal.dorda@vsb.cz.
[3] Czech Technical University in Prague / Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Konviktská 20, Prague 1, Czech Republic, e-mail: mockova@fd.cvut.cz.

a defined number of the consignments) and after finishing the service the vehicle goes back to its depot. It is more than obvious that some trips can be non-productive (so called deadheading) in the vehicle schedule. The costs that are related to vehicle deadheading may substantially devalue a solution which has been found with the Dantzig algorithm (or another methods). This devaluation which can be expressed in the total non-productively covered distance may bring an increase of the total transport costs so that the original optimal solution may lose its optimality.

In the article, an extended mathematical model of the transportation problem is presented. Contrary to the basic transportation problem, the presented mathematical model can be applied for finding an optimal solution of the transportation problem which considers all the non-productively covered distances as well. We limit ourselves to the problems for which it is given that capacity of all the vehicles equals to single consignment – that means each vehicle can carry only single consignment at the same time. A typical example of such real problem can be distribution of empty containers – each vehicle carries only single container. In addition, we assume that all the vehicles are parked in the same depot.

At first, we formulate the basic transportation problem. After doing it we present an extended modification of the transportation model which considers vehicle deadheading as well.

2 Problem formulation and state of the art

Let us consider *m* places where consignments are located – let us call them storehouses in the following text. On the other hand, *n* places where consignments are requested are given – let us call them customers.

For each storehouse $i = 1,...,m$ its capacity a_i is defined, for each customer $j = 1,...,n$ its demand b_j is known. For each possible direction of transportation between the individual storehouses and the individual customers value c_{ij} is given – the value corresponds to the distance the vehicle must cover when transporting a consignment from storehouse $i = 1,...,m$ to customer $j = 1,...,n$. All the values of c_{ij} form a unit transportation cost matrix C. Our task is to decide how many consignments should be transported from the individual storehouses to the individual customers so that the total covered distance is as minimal as possible. In the following text, the balanced transportation problem is assumed. That means it holds $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$ (the sum of the capacities of the individual storehouses – the total supply – equals to the sum of the demands of the individual customers – the total demand). In addition, it is also assumed that the capacities of the individual storehouses and the demands of the individual customers are greater than 0, that means $a_i > 0$ for $i = 1,...,m$ and $b_j > 0$ for $j = 1,...,n$.

The mathematical model of the balanced transportation problem has the following form:

$$\min \ f(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{1}$$

subject to:

$$\sum_{j=1}^{n} x_{ij} = a_i \ \text{ for } \ i = 1,...,m \tag{2}$$

$$\sum_{i=1}^{m} x_{ij} = b_j \ \text{for} \ \ j = 1,...,n \tag{3}$$

$$x_{ij} \geq 0 \ \text{ for } \ i = 1,...,m \ \text{ and } \ j = 1,...,n \tag{4}$$

Formula (1) represents the optimisation criterion, constraints (2) ensure that all the consignments found in the storehouses are transported from them (the capacity of each storehouse is depleted), constraints (3) model that each customer must be satisfied (that means each customer receives the number of the consignments that is equal to the customer's demand) and constraints (4) define the domain of definition of the variables used in the model. Model (1) – (4) will be referenced in the following text as the basic model.

The transportation problem expressed with basic model (1) – (4) can be easily solved exactly – we can employ the general simplex method [2] or the Dantzig algorithm which can be combined with ε perturbation method that prevents the solving process from obtaining degenerate solutions [1], [2]. To solve the basic transportation model many modern software tools can be also used – for example Solver add-in of Microsoft Excel or any suitable optimisation software such as Xpress-IVE.

To obtain a good initial feasible solution some heuristics can be also applied – let us mention for example the Northwest corner method, the Indexing method or the Vogel approximation method. To calculate the distance

matrix that is one of necessary pieces of input data several well-known algorithms can be applied, the algorithms are described for example in [5].

In the last paragraph of the introduction it was mentioned that the real problem which is addressed in the paper corresponds to transportation of the containers with the vehicles with the capacity which is equal to 1 container. Let us discuss some known approaches to solving the problem. Regarding Czech or Slovak publications we can mention sources [3], [4] or [7].

In [4] a combined optimisation problem of loaded container distribution, empty container distribution and vehicle deadheading is presented. The solving concept is based on forming general trips that consist of 5 phases (note that not all the phases must be included in final trips). The solving algorithm consists of two phases – the first phase deals with distribution of the empty containers and the second phase is aimed at optimisation of vehicle deadheading.

The problem of collection and distribution of the containers that is realised with the vehicles with the capacity of 2 containers is discussed in publication [3]. Publication [7] mentions some selected exact and heuristic approaches to solving distribution problems of different types of containers based on circular trips.

The above-mentioned publications also reference many foreign sources dealing with the problem.

3 Formulation and mathematical model of extended transportation problem and its solving

To formulate the extended mathematical model, original input data described in Section 2 must be extended. At first, information that the customers are supplied with the vehicles with the capacity that equals to 1 consignment must be added. A depot where all the vehicles are parked must be defined in the extended transportation problem. It is requested that the vehicles go back to the depot after finishing service of the customers.

Our task is to find a transportation plan so that all the customers are satisfied, the capacities of the individual storehouses are not exceeded and the total distance covered by all the vehicles (including deadheading) is as minimal as possible. Therefore, it is requested that after finishing optimisation calculation information how many consignments should be transported between the individual storehouses and the customers and information how the vehicles should deadhead must be provided.

Let us define notation of the extended mathematical model:

- The storehouses are labelled with indices 1,…, m – the number of the storehouses is equal to m.
- For the customers indices $m+1,...,n$ are used – the number of the customers equals to $n-m$.
- Two fictional vertices labelled with indices 0 and $n+1$ are added to the model – vertex 0 represents the depot the vehicles depart from and vertex $n+1$ the depot the vehicles return to (note that both depots could be united in a single depot because in practice both depots are usually the same).

The meaning of other symbols is the same except for the following symbols. Variables x_{ij} with indices $i = 0, m+1,...,n$ and $j = 1,...,m, n+1$ have different meaning (in comparison with the basic model). The variables model:

- How often the vehicles go empty from the depot to the storehouses (deadheading at the beginning of their schedules) – for combinations of indices $i = 0$ and $j = 1,...,m$.
- How often the vehicles go empty from the customers (after their unloading) to the storehouses (deadheading within the schedule) – for combinations of indices $i = m+1,...,n$ and $j = 1,...,m$.
- How often the vehicles go empty from the customers to the depot (deadheading at the end of their schedules) – for combinations of indices $i = m+1,...,n$ and $j = n+1$.

And finally, constant K representing the number of the vehicles that are used to realise the created transportation plan must be added to the mathematical model. As shown later in the article, an increase of K may bring worsening of the optimisation criterion.

When creating the mathematical model, a basic condition that the number of the empty vehicles arriving in each storehouse (from the depot or from any customer after unloading) is equal to the capacity of the storehouse must be satisfied. Analogously, the number of the vehicles going empty from the customer (to any storehouse or to the depot) must be equal to the demand of the customer. Both facts result from the capacity of the vehicles – as stated earlier in the article each vehicle can carry only single assignment at the same time. Each proposed solution can be depicted with a weighted directed graph; the weight of each arc corresponds to the number of the trips realised in the corresponding direction. Moreover, it must be true that the number of the arcs entering each vertex equals to the number of the arcs exiting the vertex. The rule also holds for the depots if the depots are represented with the same vertex (as assumed in the article).

The mathematical problem of the balanced transportation problem which minimises the total covered distance can be formulated in the following form:

$$\min\ f(x) = \sum_{j=1}^{m} c_{0j} x_{0j} + \sum_{i=1}^{m} \sum_{j=m+1}^{n} c_{ij} x_{ij} + \sum_{i=m+1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij} + \sum_{i=m+1}^{n} c_{in+1} x_{in+1} \tag{5}$$

subject to:

$$\sum_{j=1}^{m} x_{0j} = K \tag{6}$$

$$\sum_{i=m+1}^{n} x_{in+1} = K \tag{7}$$

$$\sum_{j=m+1}^{n} x_{ij} = a_i \ \text{ for } \ i = 1,...,m \tag{8}$$

$$\sum_{i=1}^{m} x_{ij} = b_j \ \text{ for } \ j = m+1,...,n \tag{9}$$

$$x_{0j} + \sum_{i=m+1}^{n} x_{ij} = \sum_{i=m+1}^{n} x_{ji} \ \text{ for } \ j = 1,...,m \tag{10}$$

$$\sum_{i=1}^{m} x_{ij} = \sum_{i=1}^{m} x_{ji} + x_{jn+1} \ \text{ for } \ j = m+1,...,n \tag{11}$$

$$x_{ij} \geq 0 \ \text{ for } \ i = 0,...,m \ \text{ and } \ j = m+1,...,n+1 \tag{12}$$

Formula (5) represents the optimisation criterion – the total distance covered by all the vehicles. Constraint (6) ensures that exactly K vehicles leave the depot. Constraint (7) models that the same number of the vehicles returns to the depot after serving the customers. The group of constraints (8) ensures that the capacity of any storehouse is not exceeded, the group of constraints (9) models that each customer is satisfied. The group of constraints (12) defines the domain of definition of the variables used in the mathematical model.

Let us discuss the remaining groups of constraints (10) and (11). The groups ensure that the number of loaded vehicle arrivals in the storehouses equals to the number of the loaded vehicle trips (10) and the number of the vehicle departures from the customers after unloading equals to the customers' demands (11).

Formulated mathematical model (5) – (12) can be easily solved using suitable software tools – we can employ for example Solver add-inn of Microsoft Excel or optimisation software Xpress-IVE. An important advantage of the proposed extended model is that the total number of the variables used in the model does not change significantly (in comparison with the basic model). Another advantage is that the domain of definition of the variables can be the same even if the capacity of each storehouse and the demand of each customer are integer. Therefore, no calculation obstacles should not occur when solving a problem with many storehouses and customers.

4 Calculation experiments

Functionality of the presented model was tested using a balanced transportation problem with 10 storehouses, 10 customers and 3 vehicles serving the customers. The capacities of the individual storehouses and the demands of the individual customers are given in Table 1.

i	1	2	3	4	5	6	7	8	9	10
a_i	5	4	3	12	6	8	4	12	5	15
b_j	2	7	6	14	13	12	5	5	5	5

Table 1 Input data for the storehouses and the customers

Table 2 summarises all distances between the initial depot, the individual storehouses, the individual customers and the terminal depot. Note that the distances are expressed in kilometres. If some trips are inadmissible (for example the trips from the initial depot to any customer), then the corresponding distance is equal to the prohibitive constant denoted with M.

Functionality of the model was tested in optimisation software Xpress-IVE. Results obtained by the optimisation experiment are summarised in Table 3; the values in the table represent the number of the trips that should be realised between the corresponding places. The individual places are denoted using the following abbreviations – ID is the initial depot, S_1,…, S_{10} are the storehouses, C_1,…, C_{10} are the customers and TD is the terminal depot. If no value is in a cell of the table, then no trip is planned between the corresponding places. The values

provided in the last column of the table represent the total number of the vehicles departing from the initial depot, from the individual storehouses and the customers. The values in the last row in the table correspond to the total number of the vehicles arriving in the individual storehouses and the customers and in the terminal depot.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	TD
ID	10	5	8	6	14	3	12	13	1	6	M	M	M	M	M	M	M	M	M	M	M
S_1	M	M	M	M	M	M	M	M	M	M	8	6	19	16	15	16	12	15	13	12	M
S_2	M	M	M	M	M	M	M	M	M	M	9	5	9	17	9	15	13	25	14	2	M
S_3	M	M	M	M	M	M	M	M	M	M	10	6	20	18	14	7	14	35	19	14	M
S_4	M	M	M	M	M	M	M	M	M	M	12	6	9	19	9	16	15	5	4	12	M
S_5	M	M	M	M	M	M	M	M	M	M	11	7	6	20	8	6	11	16	4	32	M
S_6	M	M	M	M	M	M	M	M	M	M	16	6	9	9	23	16	12	15	24	10	M
S_7	M	M	M	M	M	M	M	M	M	M	9	6	4	8	6	60	12	5	15	10	M
S_8	M	M	M	M	M	M	M	M	M	M	4	17	5	7	9	6	8	8	8	8	M
S_9	M	M	M	M	M	M	M	M	M	M	8	6	9	6	18	36	9	1	6	52	M
S_{10}	M	M	M	M	M	M	M	M	M	M	8	14	12	10	9	14	7	10	3	6	M
C_1	8	9	10	12	11	16	9	4	8	8	M	M	M	M	M	M	M	M	M	M	4
C_2	6	5	6	6	7	6	6	17	6	14	M	M	M	M	M	M	M	M	M	M	6
C_3	19	9	20	9	6	9	4	5	9	12	M	M	M	M	M	M	M	M	M	M	12
C_4	16	17	18	19	20	9	8	7	6	10	M	M	M	M	M	M	M	M	M	M	10
C_5	15	9	14	9	8	23	6	9	18	9	M	M	M	M	M	M	M	M	M	M	13
C_6	16	15	7	16	6	16	60	6	36	14	M	M	M	M	M	M	M	M	M	M	9
C_7	12	13	14	15	11	12	13	8	9	7	M	M	M	M	M	M	M	M	M	M	8
C_8	15	25	35	5	16	15	5	8	1	10	M	M	M	M	M	M	M	M	M	M	4
C_9	13	14	19	4	4	24	15	8	6	3	M	M	M	M	M	M	M	M	M	M	10
C_{10}	12	2	14	12	32	10	10	8	52	6	M	M	M	M	M	M	M	M	M	M	9

Table 2 The distance matrix

To / From	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	TD	Σ
ID						3																3
S_1												5										5
S_2																				4		4
S_3																3						3
S_4												2			9			1				12
S_5																6						6
S_6														8								8
S_7													4									4
S_8											2		2	5		3						12
S_9														1				4				5
S_{10}															4		5		5	1		15
C_1																					2	
C_2	5			2																		
C_3							4	2														
C_4						5		7	1	1												
C_5				10						3												
C_6			3		6			3														
C_7										5												
C_8									4													
C_9										5												
C_{10}		4								1												
Σ	5	4	3	12	6	5	4	12	5	15	2	7	6	14	13	12	5	5	5	5	3	

Table 3 The results of the experiment

The value of the objective criterion – the total covered distance including deadheading in kilometres – is 913.

As can be seen in the last column of the table, the values in rows $S_1 – S_{10}$ correspond to the capacities of the individual storehouses. Therefore, it is obvious that the numbers of the vehicles departing from the storehouses reflect their capacities. Because it is assumed that each vehicle can carry only 1 consignment at the same time, we can say that the capacities of the individual storehouses are depleted (this is one of the basic conditions of the balanced transportation problem). In the last column of rows $C_1 – C_{10}$ we can see that the values correspond to the demands of the individual customers. That means the numbers of the vehicles departing from the individual customers are equal to their demands. If the capacity of each vehicle is 1 consignment, the demands of the customers are satisfied (because the number of the vehicles arriving in the customer has to be equal to the number of the vehicles departing from the customer). Confirmation of this fact can be found in the last column of Table 3, more specifically in columns $C_1 – C_{10}$. The values in the columns express how many vehicles arrive in the individual customers. One can verify that the values are equal to the demands of the customers. The values in the last

column of the table labelled $S_1 – S_{10}$ give information how many vehicles arrive in the individual storehouses. It is obvious that the values correspond to the capacities of the individual storehouses. That means the required numbers of the vehicles arrive in the individual storehouses to be loaded. And finally, let us explain the meaning of the value which is found in the last column for the initial depot (ID) and in the last row for the terminal depot (TD). The values have to be equal to the number of the trips covered unproductively between the initial depot and the storehouses and between the customers and the terminal depot. One can see that all the vehicles start their schedule with the unproductive trip to storehouse S_6, 2 vehicles go back to the terminal depot after serving customer C_2 and the remaining vehicle after serving customer C_8.

At the end of the article let us present a piece of knowledge resulting from the calculation experiments which were devoted to research into impact of the number of the vehicles on the calculation results. Table 4 reveals how the increasing number of the vehicles *K* is changing the value of the optimisation criterion – the total covered distance – for the same transportation problem. In the table, one can see that the number of the vehicles was from 1 to 10.

K	1	2	3	4	5	6	7	8	9	10
Opt. crit. [km]	909	910	913	916	919	922	925	929	934	939

Table 4 Impact of the number of the vehicles on the optimisation criterion

5 Conclusions

In the article the optimisation problem of planning how the customers found in the transport net should be supplied from the storehouses; the customers are served by means of shuttle trips. In the article it is assumed that the capacity of the vehicles that serve the customers is equal to 1 consignment (for example 1 container). The theoretical part of the article contains the original mathematical model which minimises the total distance covered with the vehicles when serving the customers. In the article an interesting phenomenon which has been revealed within the calculation experiments with the presented mathematical model is discussed – we found out that the total covered distance increases with the increasing number of the vehicles serving the customers.

Acknowledgment

The article was supported by the research project SP2017/138 „Research into Optimisation of Transportation System Technologies“.

References

[1] Cenek, P. and Klima, V. and Janáček, J.: *Optimalizace dopravních a spojových procesů*. Vysoká škola dopravy a spojov, Žilina, 1994.
[2] Gass, S. I.: *Lineárne programovanie*. Metódy a aplikácie. ALFA, Bratislava, 1972.
[3] Janáček, J.: Metoda řešení svozu a rozvozu kontejnerů v rámci atrakčního obvodu dvojložnými tahači. In *Práce a štúdie Vysokej školy dopravy a spojov v Žiline, Séria kybernetická, Svazek. 2*.
[4] Janáček, J.: *Operační analýza II. Diskrétní programování a dopravní úlohy*. ALFA, Bratislava, 1983.
[5] Palúch, S.; Majer, T.: Review of possibilities for speed-up of distance matrix calculation. In *Proceedings of 33rd International Conference on Mathematical Methods in Economics* (Martincik, D., Ircingova, J., Janecek, P. eds.). Univerzita West Bohemia, Plzen, 2015, 596 – 601.
[6] Palúch, S. and Peško, Š. *Kvantitatívne metódy v logistike*. Žilinská univerzita, Žilina, 2006.
[7] Skýva, L.; Janáček, J.; Cenek, P. *Energeticky optimální řízení dopravních systémů*. Nakladatelství dopravy a spojů, Praha, 1983.

Subordinating Lévy processes and the measure of market activity

Tomáš Tichý[1]

Abstract. Modern approach to the analysis of financial asset prices and evolution of their returns requires utilization of stochastic processes that overcome the highest deficiency of Gaussian distribution, i.e., higher moments of the distribution are considered and/or fluctuation of parameters over time is allowed. In this contribution we focus on selected types of Lévy models defined as subordinated Brownian motion and consider the impact of various levels of market activity. Within the illustrative study we utilize FX rates market data, stressing the impact of financial crises.

Keywords: Lévy models, subordinators, market activity

JEL classification: C44, G13
AMS classification: 35, 90C15

1 Introduction

Standard approach to modelling financial instrument prices is based on the assumption of continuous-time and relatively smooth evolution of market prices. Despite that at most of markets the trading can occur at only particular periods of a day and only discrete changes in the market prices are possible (i.e., the changes cannot be infinitesimal), continuous-time diffusion-type processes has mostly been regarded as a good proxy.

On the other hand, various studies in the past have proofed that the distribution of financial instrument prices often exhibit quite high kurtosis and fat tails, which is given especially by market shocks, see, e.g. [12],[13], [10], [15], etc.

One suitable way how to distinguish between periods with different level of market activity, is the utilization of subordinated models belonging to a general family of Lévy processes. Here, the subordinator plays a role of an internal time which helps to fit the activity at a given market – for example, if there is no new information relevant for the value of a given financial instrument, there is obviously no reason to make a trade and thus the market looks like sleeping (time runs *slowly*); on the other hand, if there are many new information with potential impact on the value of a given instrument, many traders try to utilize their view on the impact and execute market orders – as a result, the price fluctuates frequently and from the standard time point of view it looks like jumping (times run very *quickly*). In this short contribution we restrict ourselves on extracting the market activity at FX rate markets, i.e., a market with high liquidity and assumed efficiency, using VG model. Related studies with further theoretical foundations can be found, e.g., in [17].

We proceed as follows – in Section 2 we review selected Lévy models with subordinator, while in Section 3 we briefly describes dependency models suitable for the market activity. Finally, in Section 4 some numerical results are provided.

2 Lévy models with subordinator

The family of Lévy models bears the name of Paul Lévy, French mathematician of the first half of 20th century and one of the founders of modern theory of stochastic processes. The history and basic principles of financial modeling via Lévy processes is studied in particular details in [8], though, more rigorous and detail treatment is provided in [4]. This family consists of such processes, whose increments are independent and stationary. Moreover, a typical feature of many processes of this family is a so called stochastic continuity – the probability of jumps at any infinitesimal time interval dt is zero.

The basic building blocks of complex Lévy models are the Wiener process and the Poisson process. It implies that any Lévy model can be decomposed into a diffusion part and a jump part. Obviously, it is sufficient that only one of these components is present, i.e., Wiener or Poisson process itself are members of the Lévy family. Moreover, it is believed that processes with infinite activity of jumps, such as *gamma process*, allows one to get a behavior similar to diffusion processes.

Modeling of financial asset prices is mostly restricted to processes that lead to positive values only, i.e., the stock price cannot drop bellow zero. It follows that we have to transform a simple Lévy model into the *exponential*

[1]VŠB – Technical University of Ostrava, Department of Finance, Sokolská 33, 701 21 Ostrava, Czech Republic, tomas.tichy@vsb.cz

Lévy models with price dynamic of the asset $\mathcal{S}(t)$ as follows:

$$\mathcal{S}(t) = \mathcal{S}_0 \exp[\mu t + X(t)].$$

A key tool for definition of advanced Lévy models is the characteristic function ϕ, so that we can avoid the problems connected with the distribution function of the random variable F_X. Let us assume a probability distribution that is infinitely divisible. Then the Lévy process is a stochastic process $X = \{X(t), t \in [0, \infty)\}$, which is zero at origin, $X(0) = 0$, if:

1. its increments are independent,
2. its increments are stationary, i.e., $X_{t+\tau} - X_t$ is of the same distribution as X_τ and does not depend on t ,
3. is stochastically continuous, $X_{t+\tau} \overset{P}{=} \lim_{\tau \to 0} X_t$.

Thus, we get the generalized properties of Wiener and also Poisson process. Moreover, it holds that the increments over time interval $\tau \geq 0$, i.e., $X_{t+\tau} - X_t$, has characteristic function $(\phi(u))^\tau$. Note also, that the third property indicates that the jumps can arise, though the probability of occurrence in any particular time instant is zero.

The cumulant of the characteristic function $\Phi(u) = \ln \phi(u)$ is called the characteristic exponent and fulfils a Lévy-Khinchin formula:

$$\Phi(u) = \imath\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{\infty} \left(\exp(\imath u x) - 1 - \imath u x \mathbb{I}_{|x|<1}\right) \nu(\mathrm{d}x). \tag{1}$$

Here, $\gamma \in \mathbb{R}$, $\sigma^2 \geq 0$ and ν is a measure on $\mathbb{R}\backslash\{0\}$ with

$$\int_{-\infty}^{\infty} (1 \wedge x^2)\nu(\mathrm{d}x) < \infty. \tag{2}$$

For a given infinitely divisible distribution we can state the triplet of Lévy characteristics, $\{\gamma, \sigma^2, \nu(\mathrm{d}x)\}$. The former two describe the drift of the process (deterministic part) and its diffusion. If it is of the form $\nu(\mathrm{d}x) = u(x)\mathrm{d}x$, we call it Lévy density. It is quite related to probability density except the fact it need not be integrable and zero at origin.

From the economic point of view, these processes can be understood as a process that evolves in a random business time. Such time is related to the arrival of new information relevant to the price of the asset we study. Since the time t, in which the Brownian motion evolves, is replaced by a stochastic process, such type of models should be regarded as a compound process and relevant probability distribution belongs to the class of mixed distributions.

The principles of modelling via internal processes, that drives the activity of the main process (i.e., subordinator), were introduced in [5] or [7]. The main idea lying behind such approach was empirically observed instability of estimated volatility and quite frequent occurrence of rare events (when compared to the Gaussian distribution), which was supported also by asymmetry properties of the returns distribution, see, for example, [10]. Overall, these empirically observed features make the usage of BS-type models based on geometric Brownian motion difficult.

If we denote $\mathcal{Z}(t; \mu, \sigma)$ as Wiener process in time t with $\mu = 1$ and $\sigma = \sqrt{t}$, i.e., $\mathcal{Z}_t = \varepsilon\sqrt{t}$, $\varepsilon \sim \mathcal{N}(0; 1)$, we can define Brownian motion $X(t; \theta, \vartheta)$ with increment θ and volatility ϑ driven by another Lévy process $\ell(t; \nu)$ with measure of variance ν simply by putting $\ell(t)$ into t. Then,

$$X_t = \theta\ell(t) + \vartheta\mathcal{Z}(\ell_t), \tag{3}$$

which can be simultaneously written as follows:

$$X_t = \theta\ell(t) + \vartheta\varepsilon\sqrt{\ell(t)}. \tag{4}$$

This relation can be interpreted in such a way that the increment dX over time change dt is of normal distribution with mean $\theta\ell_{dt}$ and variance $\vartheta^2\ell_{dt}$, since the mean of the driving process $\ell(t)$ is dt and its variance will determine the heaviness of the tails.

Quite popular subordinators are *gamma process*, leading to *VG model* (the term VG (variance gamma) was suggested since the variance of the driving process is not determined by standard time t, but *gamma time* – variance gamma) and *inverse Gaussian process*, leading to *NIG* model (*normal inverse Gaussian model*).[2]

[2]For more details, including references of particular models, see, e.g., [8].

VG model. The VG model (*variance gamma model*) is one of the most frequently applied multiparametric models of the Lévy family of processes. There exists two ways how to define it – the first one follows the definition of the Brownian motion driven by gamma process.

The probability density function of the gamma process from gamma distribution $\mathcal{G}(a, b)$ with $a = 1/\nu$ and $b = \nu$ is given by:

$$f_{\mathcal{G}}(g, t; \nu) = \frac{g^{\frac{t}{\nu}-1} \exp(-\frac{g}{\nu})}{\nu^{\frac{t}{\nu}} \Gamma(\frac{t}{\nu})}. \tag{5}$$

The process can be defined using its characteristic function as follows:

$$\phi_{\mathcal{VG}}(x, g(t; \nu); \theta, \vartheta) = \left(1 - \imath x \theta \nu + \frac{1}{2}\sigma^2 \nu x^2\right)^{-t/\nu}. \tag{6}$$

NIG model. Another popular model based on the subordinator is NIG model (*normal inverse Gaussian model*). NIG model can be defined by two ways, as well. Following the definition of the Brownian motion driven by inverse Gaussian (IG) process, i.e., process $\mathcal{I}(t; \nu)$ with drift ν, which at time $\mathcal{I} \sim \mathcal{I}G[t; \nu]$ reaches level t, we can formulate the characteristic function as follows:

$$\phi_{\mathcal{NIG}}(x; \nu, \theta, \vartheta) = \exp\left[\frac{1}{\nu} - \frac{1}{\nu}\left(\sqrt{1 + x^2\vartheta^2\nu - 2\theta\nu\imath}\right)\right], \tag{7}$$

Extension by a stochastic environment. All these models (VG, CGMY, NIG) can be easily generalized to allow a stochastic change of particular parameters. Hence, if we incorporate another internal process (time), we can obtain a random change of the overall environment. Obviously, it is a natural consequence of the relation of the stochastic volatility and the random time.

As usually, such process can be only positive. Since the real applications suggest that the process should revert back to some equilibrium, especially the Lévy models of the OU-type were studied ([2, 3]) as well as standard models of interest rates, such as the CIR model (see [9]) in [6].

Denote such process describing the stochastic environment as $y(t)$ (standard time remains t). According to CIR the increment dy within infinitesimal time interval dt can be expressed by stochastic differential equation as follows:

$$dy = k(\eta - y)dt + \lambda\sqrt{y}\varepsilon. \tag{8}$$

Here k depicts the mean reverting tendency to the equilibrium η and λ is the volatility of this random time.

Finally, we can formulate a Lévy model based on the subordinated Brownian motion within stochastic environment defined by y as follows (compare with (4)):

$$X_t = \theta\ell(Y(t)) + \vartheta\varepsilon\sqrt{\ell(Y(t))}, \tag{9}$$

where

$$Y(t) = \int_0^t y(u)\mathrm{d}u. \tag{10}$$

Generalized hyperbolic model. Generalized hyperbolic model is apparently more complicated, when compared to the preceding models, since the Lévy measure is not available in the closed form and the density function is available only for the same time scale, and even in this case its form is very complex. These features do not allow simple applications in asset price modelling. Thus, it can be useful to approximate such type of models by a (geometric) Brownian motion driven by GIG (*generalized inverse Gaussian*) subordinator.

3 Dependency and association among random variable

One of the most essential tasks of financial decision-making is the measurement of the dependency among the realizations of particular random variables. Specifically, let us consider n risky assets with gross returns[3] $z = [z_1, z_2, \ldots, z_n]^{'}$. As a consequence of the Sklar theorem ([16]) the joint distribution function is given by:

$$F_{\mathbf{z}}(\mathbf{x}) = \mathcal{C}(F_{z_1}(x_1), F_{z_2}(x_2), \ldots, F_{z_n}(x_n)), \tag{11}$$

[3] Generally, we assume the standard definition of gross return between time t and time $t+1$ of asset i, as $z_{i,t+1} = \frac{S_{i,t+1}+d_{i,[t,t+1]}}{S_{i,t}}$, where $S_{i,t}$ is the price of the i-th asset at time t and $d_{i,[t,t+1]}$ is the total amount of cash dividends between t and $t+1$.

where $F_{z_i}(x_i) = \Pr(z_i \leq x_i)$ are the marginal distribution functions and $\mathcal{C}$:$[0,1]^n \to [0,1]$ is the copula function. The copula function can therefore be defined by inverting (1):

$$\mathcal{C}(\mathbf{u}) = F_{\mathbf{z}}(F_{z_1}^{-1}(u_1), F_{z_2}^{-1}(u_2), \ldots, F_{z_n}^{-1}(u_n)). \tag{12}$$

Therefore, the dependency among particular variables is fully described by suitable copula function $\mathcal{C}$. Furthermore, the copula function can be regarded as the joint distribution function of the marginal distribution functions.

In several financial contexts it is convenient to express the dependency between random variables with a single number (more generally, for n random variables we get an n-dimensional matrix). The most widespread is *the Pearson coefficient of correlation* defined as follows:

$$\text{cor}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}, \tag{13}$$

where $\text{var}(X)$ states the variance of X and $\text{cov}(X,Y)$ the covariance between X and Y. This measure is the inner product of standardized random variables in the Hilbert $L^2 = \left\{X|E(|X|^2) < \infty\right\}$ space and it derives most of its properties from this characteristic. However, the Pearson *coefficient of correlation* is only one of the possible measures of dependency.

The most popular measures of concordance are: Kendall's tau, Spearmann's rho, Gini's gamma, and Blomqvist's beta.

The *Kendall's tau*, τ_K (also called Kendall correlation), is defined as the probability of concordance reduced by the probability of discordance:

$$\tau_K(X,Y) = \Pr\left((X_1 - X_2)(Y_1 - Y_2) > 0\right) - \Pr\left((X_1 - X_2)(Y_1 - Y_2) < 0\right), \tag{14}$$

where (X_1, Y_1) and (X_2, Y_2) are independent replications of (X, Y). Therefore,

$$\begin{aligned}\tau_K(X,Y) &= \mathbb{E}(sign((X_1 - X_2)(Y_1 - Y_2))) \\ &= \text{cor}(sign(X_1 - X_2), sign(Y_1 - Y_2)),\end{aligned}$$

where $sign(x) = 1$ if $x > 0$, $sign(x) = 0$ if $x = 0$ and $sign(x) = -1$ if $x < 0$. Clearly, Kendall's tau can be defined in terms of the copula function:

$$\tau_K(\mathcal{C}) = 4\int_0^1 \int_0^1 \mathcal{C}(u,v)d\mathcal{C}(u,v) - 1, \tag{15}$$

where $\mathcal{C}$ is the copula associated to the bivariate vector (X, Y).

The second most popular measure of concordance, *Spearman's rho*, ρ_S, is given by:

$$\begin{aligned}&\rho_S = 3\left(\Pr\left((X_1 - X_2)(Y_1 - Y_3) > 0\right) - \Pr\left((X_1 - X_2)(Y_1 - Y_3) < 0\right)\right) = \\ &= 3\mathbb{E}(sign((X_1 - X_2)(Y_1 - Y_3))) = 3\,\text{cor}\left(sign(X_1 - X_2), sign(Y_1 - Y_3)\right).\end{aligned} \tag{16}$$

where (X_1, Y_1), (X_2, Y_2) and (X_3, Y_3) are independent replications of (X, Y).

This measure is the Pearson linear correlation coefficient applied to marginal distribution functions of the random variables, i.e.:

$$\rho_S = \text{cor}\left(F_X(X), F_Y(Y)\right) = \frac{\text{cov}\left(F_X(X), F_Y(Y)\right)}{\sqrt{\text{var}\left(F_X(X)\right), \text{var}\left(F_Y(Y)\right)}}. \tag{17}$$

It follows, that it can be regarded as the correlation of copula functions:

$$\begin{aligned}\rho_S(X,Y) &= 12\int_0^1 \int_0^1 uvd\mathcal{C}(u,v) - 3 = \\ &= 12\int_0^1 \int_0^1 \mathcal{C}(u,v)dudv - 3,\end{aligned} \tag{18}$$

where $\mathcal{C}$ is the copula associated to the bivariate vector (X, Y).

Another measure used to quantify the concordance among random variables is *Gini's gamma*, γ_G. It can be defined in terms of copula functions as follows:

$$\gamma_G(\mathcal{C}) = 4\left[\int_0^1 \mathcal{C}(u, 1-u)du - \int_0^1 [u - \mathcal{C}(u,u)]du\right], \tag{19}$$

where $\mathcal{C}$ is the copula associated to the bivariate vector (X, Y). Its sample estimation is given by ranks p_i and q_i of random variables X and Y, respectively:

$$\gamma_G(X,Y) = \frac{1}{\lfloor n^2/2 \rfloor}\left[\sum_{i=1}^{n} |p_i + q_i - n - 1| - \sum_{i=1}^{n} |p_i - q_i|\right]. \tag{20}$$

Finally, we should mention *Blomqvist beta*, β_B, defined as follows:

$$\begin{aligned}\beta_B(X,Y) = \Pr[(X-\tilde{x})(Y-\tilde{y}) > 0] - \Pr[(X-\tilde{x})(Y-\tilde{y}) < 0] = \\ = \mathbb{E}(sign((X-\tilde{x})(Y-\tilde{y}))),\end{aligned} \tag{21}$$

where $\tilde{x}$ and $\tilde{y}$ are the medians of some given continuous random variables *X* and *Y*, respectively. With certain simplifications, this measure may also be rewritten in terms of copula functions:

$$\beta_B(\mathcal{C}) = 4\mathcal{C}\left(\frac{1}{2}, \frac{1}{2}\right) - 1. \tag{22}$$

The proof that all these measures are really measures of concordance can be found, for example, in Nelsen (2006).

4 Numerical illustration

Let us assume FX rates with respect to EUR over 2005 to 2011 (we select this period in order to stress the change of the market activity proxy during the subprime crises). We consider large selection of FX rates, starting with Czech Koruna (CZK), which has been supposed to be a currency with quite stable behaviour, across rather stabilized US dollar (USD) and UK pound (GBP), little more risk Hungarian Forint (HUF) and Polish Zloty (PLN) and two examples from eastern Asia, Korean Won (KRW) and Indonesian Rupiah (IDR).

Extracted proxy of market activity in terms of subordinated process on 4-year estimation window, and thus, the flow of information, is apparent from Figure 1. It is quite obvious that the activity with IDR rather differs from other currencies, especially as concerns post-crises behaviour and some high activity observed at the beginning of the monitored period. It also seems that the activity at USD and GBP market started well in advance (comparing with other FX rates). Moreover, the USD activity remained rather high even in post-crises period. Finally, also HUF and PLN show some differences and really high activity even outside crises period, especially as concerns HUF.

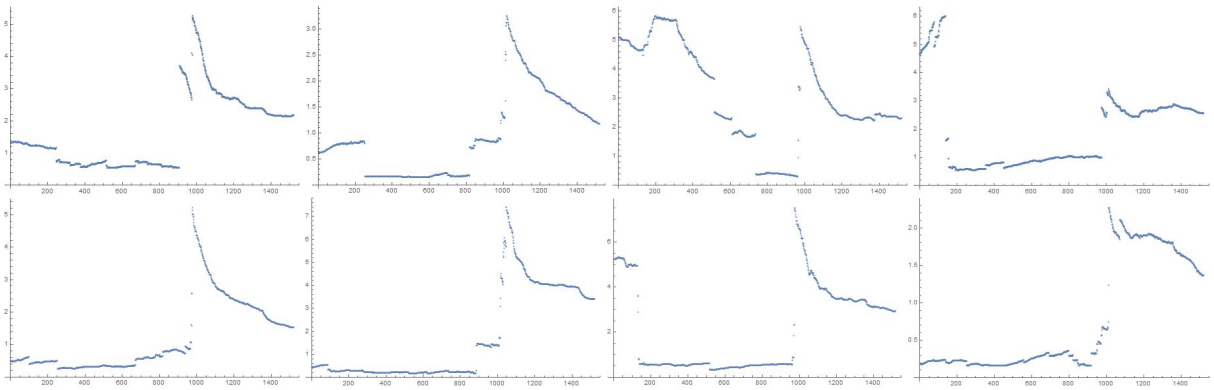


Figure 1 Estimation of market activity via VG model and its Lévy density over 2003-2012, including the financial crises impact for CZK, GBP, HUF, IDR, JPY, KRW, PLN, and USD

Similar results can be obtained for further subordinated models (we do not report them due to lack of space). As an illustration, however, Table 1 shows Kendall tau obtained for VG model and NIG model, respectively.

VG								NIG							
1	0.63	0.08	0.32	0.59	0.74	0.61	0.55	1	0.63	0.16	0.32	0.61	0.74	0.60	0.55
0.63	1	0.01	0.45	0.74	0.77	0.53	0.62	0.63	1	0.10	0.44	0.74	0.76	0.56	0.60
0.08	0.01	1	-0.10	-0.12	0.03	0.21	-0.13	0.16	0.10	1	-0.07	-0.03	0.09	0.23	-0.07
0.32	0.45	-0.10	1	0.51	0.43	0.54	0.44	0.32	0.44	-0.07	1	0.51	0.43	0.58	0.45
0.59	0.74	-0.12	0.51	1	0.69	0.54	0.69	0.61	0.74	-0.03	0.51	1	0.71	0.60	0.68
0.74	0.77	0.03	0.43	0.69	1	0.60	0.62	0.74	0.76	0.09	0.43	0.71	1	0.60	0.63
0.61	0.53	0.21	0.54	0.54	0.60	1	0.37	0.60	0.56	0.23	0.58	0.60	0.60	1	0.40
0.55	0.62	-0.13	0.44	0.69	0.62	0.37	1	0.55	0.60	-0.07	0.45	0.68	0.63	0.40	1

Table 1 Comparison of Kendall tau of market activity in 4-year time windows

5 Conclusion

We have presented market activity approximation in terms of subordinated Lévy models and in particular, results obtained using VG model have been illustrated also graphically. However, other models brings relatively similar results.

The results can be further analysed as concerns their dependency and interconnections especially in tails (i.e., what are the similarities during extreme periods). As concerns equity markets, the market activity proxy might be useful as an alarm clock suggesting conservative investments as document in [1] and further extended in [11].

Acknowledgements

The research was supported through the Czech Science Foundation (GAČR) under project 16-09541S and SP2017/32, an SGS research project of VŠB-TU Ostrava. The support is greatly acknowledged.

References

[1] Giacometti, R., Ortobelli, S., Tichý, T.: Portfolio selection with uncertainty measures consistent with additive shifts. *Prague Economic Papers* 24 (1): 3–16, 2015.

[2] Barndorff-Nielsen, O.E., Shephard, N.: Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial econometrics. *Journal of the Royal Statistical Society*, Series B 63: 167–241, 2001.

[3] Barndorff-Nielsen, O.E., Shephard, N.: Impact of jumps on returns and realised variances: econometric analysis of time-deformed Lévy processes. *Journal of Econometrics* **131**(1–2): 217–252, 2006.

[4] Bertoin, J.: *Lévy Processes.* Cambridge: Cambridge University Press, 1996.

[5] Bochner, S.: Diffusion equation and stochastic processes. *Proceedings of the National Academy of Science of the United States of America* **85**: 369–370, 1949

[6] Carr, P., Geman, H., Madan, D.B., Yor, M.: Stochastic Volatility for Lévy Processes. *Mathematical Finance* **13**: 345–382, 2003.

[7] Clark, P.K.: A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica* **41**: 135–156, 1973

[8] Cont, R., Tantkov, P.: *Financial Modelling with Jump Processes*, 2nd ed. Boca Raton: Chapman & Hall/CRC press, 2017.

[9] Cox, J.C., Ingersoll, J., Ross, S.A.: A theory of the term structure of interest rates. *Econometrica* **53**: 385–407, 1985.

[10] Fama, E.: The behavior of stock market prices. *Journal of Business* **38**, 34–105, 1965.

[11] Kouaissah, N., Ortobelli, S., Tichý, T. On the use of the conditional expectation in portfolio selection problems. *Annals of Operations Research*, to appear in 2018.

[12] Mandelbrot, B.: New methods in statistical economics. *Journal of Political Economy* 71, 421–440, 1963a.

[13] Mandelbrot, B.: The variation of certain speculative prices. *Journal of Business* **26**, 394–419, 1963b.

[14] Ortobelli, S., Tichý, T.: On the impact of semidenite positive correlation measures in portfolio theory. *Annals of Operations Research* **235**: 625–652, 2015.

[15] Rachev, S., Mittnik, S.: *Stable Models in Finance.* Wiley, Chichester, 2000.

[16] Sklar, A.: Fonctions de rpartition a n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**: 229–231, 1959.

[17] Swishchuk, A.: *Change of Time Methods in Quantitative Finance*. Cham: Springer, 2016.

[18] Tichý, T.: *Lévy Processes in Finance: Selected applications with theoretical background.* SAEI, vol. 9. VŠB-TU Ostrava, Ostrava, 2011.

Measuring Co-movements Based on Quantile Regression Using High-frequency Information: Asymmetric Tail Dependence

Petra Tomanová[1]

Abstract. The aim of this paper is to investigate whether contagion was present in sector equity markets formed from S&P constituents during crises between 1998 – 2016. The paper deals with the problem of spurious correlation measures due to the heteroskedasticity which might lead to improper inferences about the presence of contagion. The problem is solved by incorporating information about time-varying volatility into the methodology based on estimation of conditional probability of co-movements between equity returns.
The proposed methodology is split into three stages. First the integrated variance is estimated using various approaches such as realized kernel and the pre-averaging approach which incorporates data sampled in high-frequency. Second the time-varying quantiles are modeled as conditional autoregressive quantiles using estimated variance. Finally, the conditional probability of co-movement is estimated based on co-exceedance of corresponding quantiles of random variables. The model is then used for investigating asymmetric behavior of contagion during crisis on equity returns of S&P 500 index. For some sectors the results report a significant increase in sector equity return comovements in bear markets during the crisis of period between years 1998 and 2016 and insignificant changes in bull markets for all sector indices.

Keywords: CAViaR, co-movement, conditional quantile, high-frequency data.

JEL classification: C22, C32, G15
AMS classification: 91G70

1 Introduction

Analyzing dependencies among equity markets is an important issue for market participants in particular because a loss in one market can be accompanied by a loss in another market. Hence an underestimating of dependencies among financial markets may cause underestimating of potential losses. But how to estimate the dependencies properly especially during financial crisis? Forbes et al. [6] and Longin et al. [8] among others show that the solution is not a straightforward at all and misleading results have often been reported in the past. The goal of this paper is twofold: First, to investigate the possibly asymmetric behavior of co-movements in the tranquil and crisis periods and second, whether incorporating information from high-frequency data in form of realized volatility would change the results.

For the purpose of showing possible solution to these problems the paper follows and extends the methodology of time-varying regression quantiles introduced by Cappiello et al. [4] in Section 2 since it exhibits many suitable properties. It allows to identify and measure asymmetries in comovement in the upper and lower tails of the distributions and no assumptions neither on the joint distributions nor on the marginal distributions of variables are needed since the regression quantile is a semi-parametric technique. It is robust to outliers, however, it is not necessary robust to heteroskedasticity and thus the problem of spurious correlation might occur. In Section 4 we investigate whether it might be problem when analyzing sector indices formed from S&P constituents. Section 5 concludes.

2 Estimating comovements

The methodology of measuring comovements using high-frequency information could be split into four parts. First the realized volatility is estimated which is used in the second part for time-varying quantiles modeling by quantile regression. Then average probabilities of comovements are estimated by OLS regression. Finally, the changes in financial comovements between tranquil and crisis periods are tested.

[1]University of Economics, Department of Econometrics, W. Churchill sq. 4, Prague 3, Czech Republic, petra.tomanova@vse.cz

2.1 Realized variance

Hypothesis of this paper is that the volatility process of prices should be considered for modeling time-varying quantiles and probability of comovements estimation, especially when data sample consist of both tranquil and turbulent periods. We employ the realized variance (RV) since Andersen et al. [1] among others showed that the RV provides a satisfactory accurate measure of volatility.

RV is the sum of squared returns through given time period. In this paper, as we are interested in a measure of daily price variation, the RV is computed as the sum of n squared intraday returns for a particular day:

$$RV_t^{(N)} = \sum_{r=1}^{N} (p_{t_r} - p_{t_{r-1}})^2$$

where p_{t_r} is r-th price observed at day t. Under particular assumptions the RV consistently estimates the quadratic variation of the price process [2]. However, in practice it is not commonly the case. Especially due to the fact that the p_t is not usually equal to efficient price p_t^* but it is contaminated by the so called microstructure noise ε_t, e.g. $p_t = p_t^* + \varepsilon_t$. Thus robust techniques have to be used when one is interested in price variation without microstructure noise. In this paper both the realized kernel estimator of [3] and pre-averaging estimator of [7] is considered.

2.2 Time-varying quantiles

Let $\{y_{it} : i = 1, \ldots, n\}$ be a set-up with n random variables at time $t \in [1, T]$. Then the conditional time-varying quantile q_{ijt} for a given j-th confidence level $\theta_j \in (0, 1)$ and for a random variable y_{it} conditional on $\mathcal{F}_{t-1}$ is defined as

$$\Pr[y_{it} \leq q_{ijt} | \mathcal{F}_{t-1}] = \theta_j.$$

Individual quantiles are modeled in accordance with the methodology of Engle et al. [5]. The authors propose a conditional autoregressive value at risk by the regression quantiles (CAViaR) models which were primarily constructed to estimate the value at risk (VaR). Since VaR is defined as a particular quantile of future portfolio values conditional on current available information, $\Pr[y_t < VaR_t | \mathcal{F}_{t-1}] = \theta$, the problem of finding the VaR_t is equivalent to estimating the time-varying conditional quantile. A generic CAViaR model can be specified as

$$q_{it}(\boldsymbol{\beta}_{ij}) = \beta_{0,ij} + \sum_{k=1}^{v} \beta_{k,ij} q_{i,t-k}(\boldsymbol{\beta}_{ij}) + \sum_{l=1}^{w} \beta_{v+l,ij} f(\mathbf{x}_{i,t-l}),$$

where $f(\mathbf{x}_{i,t-l})$ is a function of a finite number of lagged observable variables of order l and $q_{it}(\boldsymbol{\beta}_{ij})$ is an empirical specification for the q_{ijt} where $\boldsymbol{\beta}_{ij}$ is the p-vector of parameters. Autoregressive terms $\beta_{k,ij} q_{i,t-k}(\boldsymbol{\beta}_{ij})$, $k = 1, \ldots, v$, ensure that the quantile changes smoothly over time and function $f(\mathbf{x}_{i,t-l})$ links $q_{it}(\boldsymbol{\beta}_{ij})$ to observable variables that belong to the information set.

Vector of parameters is estimated using the quantile regression, thus an estimator $\hat{\boldsymbol{\beta}}_{ij,T}$ of an unknown vector of parameters $\boldsymbol{\beta}_{ij}$ is defined as

$$\min_{\boldsymbol{\beta}_{ij}} \frac{1}{T} \sum_{t=1}^{T} (\theta_j - \mathbb{1}[y_{it} \leq q_{it}(\boldsymbol{\beta}_{ij})])(y_{it} - q_{it}(\boldsymbol{\beta}_{ij})),$$

where $\mathbb{1}[e \leq 0]$ is equal to one when $e \leq 0$ and zero otherwise. Once the estimates $\hat{\boldsymbol{\beta}}_{ij,t}$ is obtained, conditional quantiles $q_{it}(\hat{\boldsymbol{\beta}}_{ij})$ can be computed.

2.3 Average probability of comovement

The conditional probability of comovement is estimated with accordance to Cappiello et al. [4]. The idea behind the methodology lays in estimation of a probability that a random variable y_{1t} falls below a conditional quantile q_{1jt} given that other random variable y_{2t} is also bellow its corresponding quantile q_{2jt}.

When we generalize the problem for n-tuple of random variables $\{y_{it} : i = 1, \ldots, n\}$ then an average probability that all random variables $\{y_{it} : i = 1, \ldots, n\}$ fall simultaneously below their quantiles over a given time period is defined as $\overline{F}_{jT} \equiv T^{-1} \sum_{t=1}^{T} F_t(q_{1jt}, \ldots, q_{njt})$. In this setting $F_t(q_{1jt}, \ldots, q_{njt})$ is a cumulative distribution function for a set of quantiles $(q_{1jt}, \ldots, q_{njt})$. Conditional quantiles are estimated by regression quantiles in accordance to Section 2.2.

For the purpose of obtaining the average probability of comovement $\overline{F}_{jT}$, n indicator variables $I_{it}(\boldsymbol{\beta}_{ij}) \equiv \mathbb{1}[y_{it} \leq q_{it}(\boldsymbol{\beta}_{ij})]$ are constructed. Then the following linear regression model can be defined and estimated:

$$\prod_{i=1}^{n} I_{it}(\hat{\boldsymbol{\beta}}_{ij,T}) = \boldsymbol{W}_t \boldsymbol{\alpha}_j + \epsilon_t, \;\; j = 1, \ldots, m, \tag{1}$$

where $\boldsymbol{W}_t \equiv [1 \; \boldsymbol{S}_t \; \boldsymbol{X}_t]$, $\boldsymbol{\alpha}_j$ is an s-vector of unknown parameters, $\boldsymbol{S}_t$ denotes an s_1 row vector of time dummies, $\boldsymbol{X}_t$ denotes an s_2 row vector of exogenous variables and $s = 1 + s_1 + s_2$.

Let $\hat{\boldsymbol{\alpha}}_{j,T}$ be the OLS estimator of (1), $\hat{\alpha}_{lj,T}$ be the $(l+1)$-th element of $\hat{\boldsymbol{\alpha}}_{j,T}$ for $l = 0, 1, \ldots, s_1$ and $S_{l,t}$ represent the l-th element of $\boldsymbol{S}_t$. Row vector $\boldsymbol{S}_t^{(-l)}$ denotes the vector $\boldsymbol{S}_t$ from which the l-th element is removed. Then dummies $\boldsymbol{S}_t$ are defined as $\{S_{l,t} = 1, \; \boldsymbol{S}_t^{(-l)} = \boldsymbol{0}\}_{t=1}^{T}$, where $\boldsymbol{0}$ is a zero vector of required dimension, in accordance with paper of Cappiello et al. [4].

Cappiello et al. [4] provided a set of conditions under which the estimated intercept $\hat{\alpha}_{0j,T}$ of (1) converges in probability to the average probability of comovement in the period of $t \in \{t : \boldsymbol{S}_t = \boldsymbol{0}\}$ and the sum of estimated parameters $\hat{\alpha}_{0j,T} + \hat{\alpha}_{lj,T}$ converges to the average probability of comovement in the period corresponding to the dummy.

2.4 Testing for contagion

For the purpose of this study the realization of random variable y_{it} is a return of an equity index for i-th country at time t and the dummy variable, $\boldsymbol{S}_t = [S_t]$, from the regression (1) identifies crisis times.

When

$$\begin{aligned} F_t^-(\theta_j) &\equiv \theta_j^{-1} \mathrm{Pr}(y_{1t} \leq q_{1t}(\boldsymbol{\beta}_{1j}), \ldots, y_{nt} \leq q_{nt}(\boldsymbol{\beta}_{nj})), \\ F_t^+(\theta_j) &\equiv (1-\theta_j)^{-1} \mathrm{Pr}(y_{1t} \geq q_{1t}(\boldsymbol{\beta}_{1j}), \ldots, y_{nt} \geq q_{nt}(\boldsymbol{\beta}_{nj})) \end{aligned}$$

for the given confidence level θ_j then the likelihood $p_t(\theta_j)$ of a tail event at time t for any subset Γ_t of random variables $\{y_{it} : i = 1, \ldots, n\}$ given that a tail event occurred for $\{y_{it} : i = 1, \ldots, n \; \wedge \; y_{it} \notin \Gamma_t\}$ is defined as

$$p_t(\theta_j) \equiv \begin{cases} F_t^-(\theta_j) & \text{if } \theta_j \leq 0.5 \\ F_t^+(\theta_j) & \text{if } \theta_j > 0.5 \end{cases}.$$

The probability of comovement in tranquil times $\overline{p}_0(\theta_j)$ and the probability of comovement in crisis times $\overline{p}_1(\theta_j)$ is defined as

$$\begin{aligned} \overline{p}_0(\theta_j) &\equiv C_0^{-1} \sum_{t \in \{t : \boldsymbol{S}_t = \boldsymbol{0}\}} p_t(\theta_j), \\ \overline{p}_1(\theta_j) &\equiv C_1^{-1} \sum_{t \in \left\{t : S_t = 1, \boldsymbol{S}_t^{(-1)} = \boldsymbol{0}\right\}} p_t(\theta_j), \end{aligned}$$

where C_0 and C_1 is a number of observations during tranquil and crisis period respectively. Cappiello et al. [4] applied the following rigorous joint test for an increase in comovements:

$$\hat{\delta}(\underline{\theta}, \overline{\theta}) = (\#\theta)^{-1} \sum_{\theta_j \in [\underline{\theta}, \overline{\theta}]} [\overline{p}_1(\theta_j) - \overline{p}_0(\theta_j)] = (\#\theta)^{-1} \sum_{\theta_j \in [\underline{\theta}, \overline{\theta}]} \hat{\alpha}_{1,j},$$

where $\#\theta$ is a number of addends in the sum and $\hat{\alpha}_{1,j}$ is the OLS estimate of $\alpha_{1,j}$ in (1) which denotes parameter belonging to S_t. When the null hypothesis $\mathrm{H}_0 : \hat{\delta}(\underline{\theta}, \overline{\theta}) = 0$ is rejected in favor of the alternative hypothesis $\mathrm{H}_1 : \hat{\delta}(\underline{\theta}, \overline{\theta}) \neq 0$ we conclude that the comovements do change between tranquil and crisis periods.

3 Data

In this paper equity returns of S&P 500 constituents are analyzed. The sample includes intraday prices (1 minute data) of 500 large companies from January 2, 1998 to November 29, 2016 having common stock listed on the NYSE or NASDAQ. The sample period spans over 4 759 days which results in 4 758 continuously compounded daily log returns for each of the companies. To ensure the synchronicity we consider only returns observed between trading hours (9:30 – 16:00). Then the 11 volume-weighted index is computed for the sectors: 'Health Care' (HeC), 'Industrials' (Ind), 'Consumer Discretionary' (CoD), 'Information Technology' (InT), 'Consumer Staples'

(CoS), 'Utilities' (Uti), 'Financials' (Fin), 'Real Estate' (ReE), 'Materials' (Mat), 'Energy' (Ene) and 'Telecommunications Services' (TeS). Each sector index consist of ten companies belonging to the sector with the highest market capitalizations.

Descriptive statistics and sample characteristics for each country were investigated. The presence of financial crises in the sample results in extreme returns and strong fluctuation. Common characteristics of daily returns indicate leptokurtic distribution and the null hypothesis of the Jarque-Bera test, i.e of normally distributed returns, is rejected at the 1% significance level. The presence of autocorrelation is confirmed by the Ljung-Box test and the presence of a unit root in the return series is readily rejected by the Augmented Dickey-Fuller test at the 1% significance level. Hence the methodology is suitable for analyzing the data sample.

We determine 5 crisis periods according to Cappiello et al. [4]: Russian crises – August 3, 1998 to December 31, 1998; Argentinean crisis – March 26, 2001 to May 15, 2001; the US sub-prime crisis – February 15, 2007 to March 30, 2007; Lehman bankruptcy – September 1, 2008 to October 31, 2008; turbulences in the euro area – August 1, 2011 to September 30, 2011. The consequent period of September 4, 2012 – November 29, 2016 represents rather a tranquil period of time, thus additional crisis periods are not defined. The crisis sample includes 261 daily equity returns.

A clear evidence of higher correlations over turbulent times than over tranquil times is found in the data for sector indices. The highest increase of correlation is indicated between the 'Industrials' sector and 'Information Technology' sector, for which the excess correlation between tranquil and crisis periods is equal to 0.50. The effect is not present in returns of sectors 'Health Care' when pairing with 'Financials', 'Real Estate' and 'Materials'. For all bivariate combinations of 11 sectors the average correlation is approximately 0.14 and 0.27 over tranquil days and over days of turbulence respectively. Forbes et al. [6] pointed out that such differences are caused by the heteroskedasticity issue rather than by an increased dependence among country-specific returns during the crises. in this paper an improper inference due to the spurious correlation is avoided since we control for volatility when the average probability of co-movement is estimated as well as in modeling conditional quantiles.

4 Results

Fist the realized kernels and realized volatility using pre-averaging approach are estimated. Since the values are similar from the two approaches, only results using realized kernels are reported.

The following non-linear CAViaR model specification is selected:

$$q_{it}(\boldsymbol{\beta}_{ij}) = \beta_{0,ij} + \beta_{1,ij} q_{i,t-1}(\boldsymbol{\beta}_{ij}) + \beta_{2,ij} y_{i,t-1} - \beta_{1,ij}\beta_{2,ij} y_{i,t-2} + \beta_{3,ij}|y_{i,t-1}| + \beta_{4,ij} S_t + \beta_{5,ij} RK_t^{(N)}, \quad (2)$$

where $RK_t^{(N)}$ is realized kernel computed using $N = 391$ intraday returns. Then the model (1) is estimated where $\boldsymbol{X}_t = \left[RK_t^{(N)}\right]$. The selected CAViaR model is estimated for 11 different sector idices and 19 quantiles ranging from 0.05 to 0.95 with step equal to 0.05.

All possible variations of 11 sector pairs are analyzed resulting in 55 sector pairs for the period of January 2, 1998 to November 29, 2016. Four null hypotheses $\hat{\delta}(\underline{\theta}, \overline{\theta}) = 0$ of different intervals are tested for each sector pair twice – first the volatility is not considered in equations (1) and (2), and secondly the realized kernel is included as explanatory variable. The results are reported in the Figure 1 and 2 where the t statistics are visualized for the case without and with realized kernel respectively. The values are computed from the average of $\hat{\alpha}_{1,j}$ over θ_j, i.e. $\hat{\delta}(\underline{\theta}, \overline{\theta}) = (\#\theta)^{-1} \sum_{\theta_j \in [\underline{\theta}, \overline{\theta}]} \hat{\alpha}_{1,j}$, and the associated standard errors. The t statistic provide a joint test for changes in co-movements between two countries. The blue color indicates significant changes in comovements between tranquil and crisis periods and the red color indicates insignificant ones. Upper matrices represents results for the upper tail of distributions: left matrix represents results for $\hat{\delta}(0.95, 0.75)$ and the right matrix for $\hat{\delta}(0.50, 0.75)$. And similarly, lower matrix represents results for the lower tail of distribution: left matrix represents results for $\hat{\delta}(0.25, 0.50)$ and the right matrix for $\hat{\delta}(0.05, 0.25)$.

First striking evidence from Figures 1 and 2 is that the inclusion of realized kernel into estimation procedure of conditional probability of co-movements changes the results. The difference occurs mainly in the lower tail of distribution for quantiles in the interval of (0.05,0.25): based on model without the realized kernels we observe that contagion is present between 83.6% of sector pairs, however when we include the realized kernels into model than we conclude that the contagion is present only in 43.6% of cases. Using the model with realized kernels the null hypothesis for $\hat{\delta}(0.75, 0.95)$ is rejected for all sector pairs (contagion is not presented). Only for 'Materials' – 'Energy' and 'Information Technology' – 'Consumer Staples' pairs the null hypothesis for $\hat{\delta}(0.50, 0.75)$ is not rejected, for the rest 53 sector pairs the contagion is not presented. The contagion in the lower tail of distribution is significant mainly for pairs formed with 'Information Technology' sector.

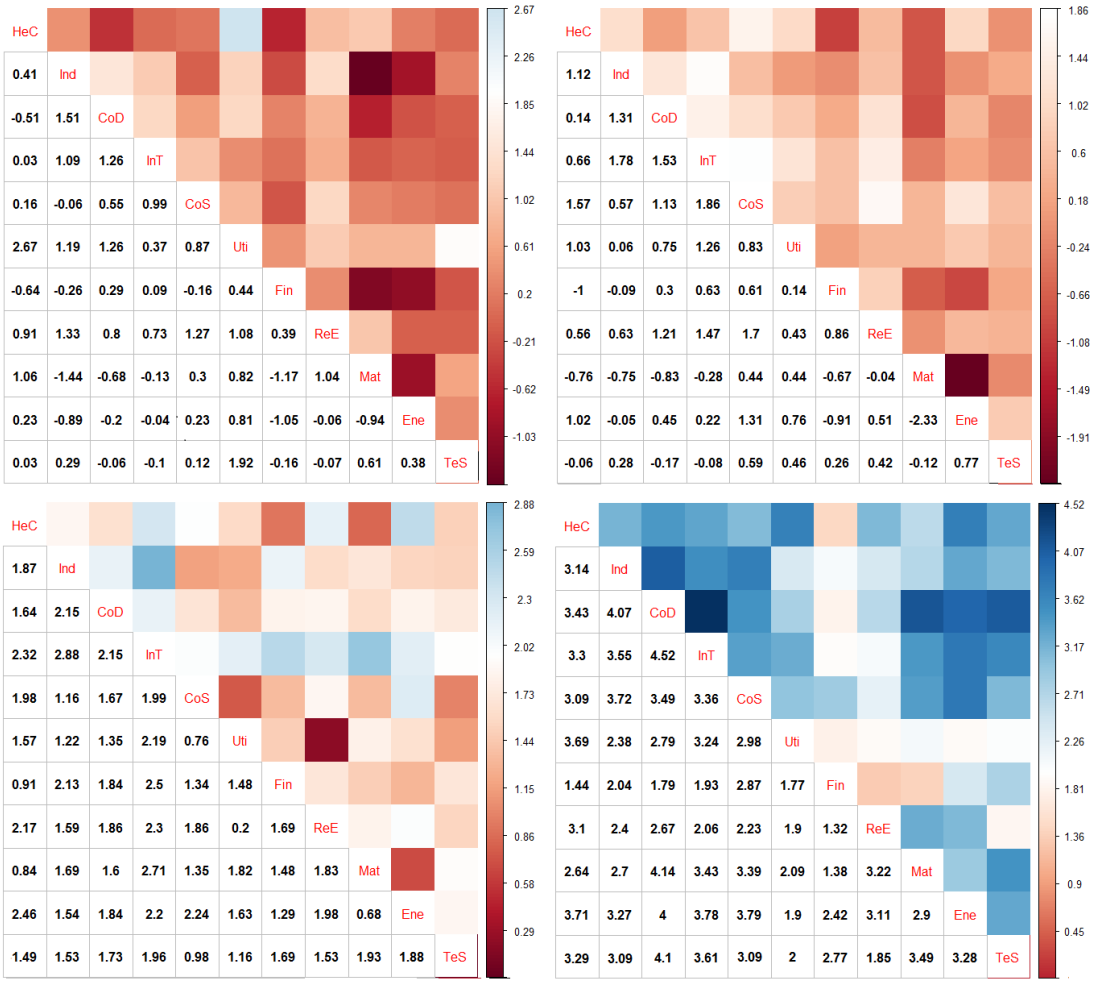


Figure 1 Results from the joint test for changes in comovements (t statistic) – volatility is not considered

Overall, it is evident that volatility should be considered when testing for contagion. Secondly, these results confirmed that the distributions of returns for sectors are characterized by asymmetries. These asymmetries cannot be detected by a simple correlations or estimates of adjusted correlation coefficients of Forbes et al. [6].

5 Conclusion

In this paper the methodology of measuring changes in comovements was extended with high-frequency information and applied to 11 sector equity markets. Between several sectors indices especially between pairs formed with 'Information Technology' the results show that comovements in equity returns across sectors tend to increase significantly in turbulent times against tranquil times in the lower tail of the distribution for the period of January 1, 1988 to April 30, 2015. However, for vast majority of analyzed sectors the differences are not significant in the upper tail of the distribution. For some sector this asymmetric dependence documents a significant increase in sector equity return comovements in bear markets during the crisis of period between years 1998 and 2016 and insignificant changes in bull markets for all sector indices. The main finding is that it is important to control for volatility since the results change in 28% cases when analyzing bear markets.

Acknowledgements

This work was supported by the research grant VŠE IGA F4/63/2016, Faculty of Informatics and Statistics, University of Economics, Prague. This support is gratefully acknowledged.

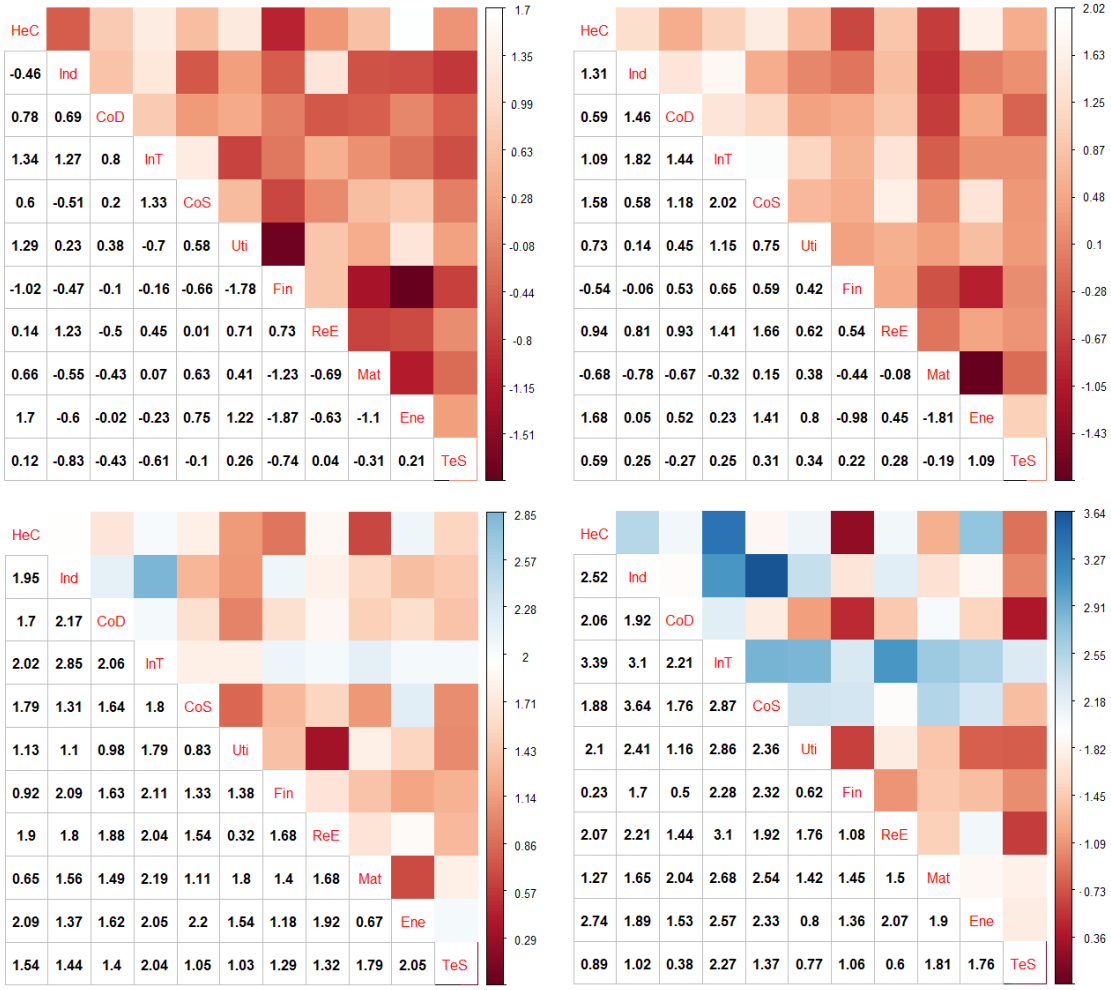


Figure 2 Results from the joint test for changes in comovements (t statistic) – realized kernels are included

References

[1] Andersen, T. G., and Bollerslev, T.: Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International economic review* (1998), 885-905.

[2] Barndorff-Nielsen, O. E.: Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64** (2002), 253-280.

[3] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N.: Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76** (2008), 1481-1536.

[4] Cappiello, L., Kadareja, A., and Manganelli, S.: Measuring comovements by regression quantiles, *Journal of Financial and Quantitative Analysis* **45** (2010), 473–502.

[5] Engle, R. F., and Manganelli, S.: CAViaR: Conditional autoregressive value at risk by regression quantiles, *Journal of Business & Economic Statistics* **22** (2004), 367–381.

[6] Forbes, K. J., and Rigobon, R.: No contagion, only interdependence: measuring stock market comovements, *The journal of Finance* **57** (2002), 2223–2261.

[7] Jacod, J., Li Y., Mykland, P. A., Podolskij, M., and Vetter, M.: Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications* **119** (2009), 2249-2276.

[8] Longin, F., and Solnik, B.: Extreme correlation of international equity markets, *The journal of finance* **56** (2001), 649–676.

Systemic Risk and Community Structure in the European Banking System

Gabriele Torri[1]

Abstract. Financial contagion and systemic risk have become increasingly relevant after the financial crisis in 2008. Network theory is a powerful framework for the analysis of these phenomena and is becoming a standard tool in the literature. This paper investigates the properties of the European banking system, focusing on the community structure of the network to identify the potential channels for the propagation of financial distress. The network structure is estimated from the sparse partial correlation of CDS spreads using *tlasso*, a robust technique that induces sparsity in the network. The optimal community structure is then estimated by a procedure that maximises modularity. The analysis shows that, despite the high level of internationalization of the financial system, it exist a clear community structure that mirrors the geographical location of the banks. Finally, a decomposition of *strength centrality* based on the estimated community structure is provided. Such decomposition represents a useful and easy-to-implement tool to monitor the exposure to financial contagion, integrating the traditional risk management tools.

Keywords: Community detection, systemic risk, network theory

JEL classification: C44
AMS classification: 90C15

1 Introduction

Systemic risk is an increasingly relevant issue, especially after the 2008 financial crisis. Given the complexity of the phenomenon, it is difficult to provide a unique definition: concerning the banking sector, some authors put particular emphasis on the effect of macroeconomic shocks on economic fundamentals (see for instance [9]), while other authors focus on the interaction between the public and the financial sectors, stressing the spillover effects of the crisis from the financial system to governments' balances [2]. Most of the works however, stress the idea of shocks affecting financial institutions and/or markets, that can propagate to the entire system [15]. In this context, network theory has become in recent years a fundamental tool for the analysis of systemic risk, capable of describing the structure of the network and to model the diffusion of distress.

Here we focus on the issue of estimating and analysing the structural properties of the banking system network. Such network is often identified as the bilateral exposures on the interbank market [8], that however are often non disclosed and they have to be inferred in absence of bilateral data. A strand of literature reconstructs the network using the total exposures of each bank towards the entire banking system through statistical techniques such as maximum entropy [12]. An alternative approach is to consider the co-movement of time series as proxies for the banks' interdependencies, and to use them to infer the network structure. (e.g. [4] and [14]). We follow this strand of literature, estimating the network from the partial correlations of banks' Credit Default Swap (CDS) spreads. In particular we use the *tlasso* algorithm, an efficient procedure to estimate the sparse partial correlation structure under the assumption of a multivariate t-Student distribution [6]. *Tlasso* can be considered an extension of *glasso*[10], an algorithm that relies on the normality assumption. Compared to the latter, *tlasso* is more appropriate for data with fatter tails and is more robust to misspecification and outliers.

After the estimation of the network, we analyse its structural properties. The literature is mostly focused on national banking systems, that typically present a highly sparse and tiered structure [12]. International banking systems are less studied; the available literature describes a more complex structure compared to national systems (see [1] and [5]). In this paper we focus on a particular structural properties of a network: the presence of a community structure. We identify the communities using the algorithm in [13], that maximises modularity to get the optimal partition, selecting also the optimal number of communities. Thanks to a rewiring procedure, we can also test the significance of the community structure that we identify against an appropriate null model. We find that the European banking system is characterized by a strong and stable community structure, and that this structure is largely overlapping to geographical divisions. In the final part we discuss the policy implication of our results and we provide a simple application that highlights the usefulness of community detection in the regulatory framework.

[1]University of Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy, gabriele.torri@unibg.it

2 Methods

2.1 Network Estimation

We briefly present the topic of partial correlation networks in the context of Gaussian graphical models, for more details on the mathematical derivation the reader is referred to [11].

Let $X \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a multivariate Gaussian distribution. This distribution can be associated to an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the nodes in $\mathcal{V}$ correspond to each element of X, the edges $\mathcal{E}$ consist of the pairs of random variables with non-zero partial correlations: $\mathcal{E} = \{(i,j) \in \mathcal{V} \times \mathcal{V} | \rho_{ij} \neq 0\}$ and the edge weights consist in the corresponding partial correlations $\rho_{i,j}$. Partial correlations ρ_{ij} are the linear dependence between two variables conditional on all other variables, and they can be related to the variance-covariance matrix of X. If we define the matrix $\boldsymbol{\Omega} = (Cov[X])^{-1}$ (hereafter defined the *precision matrix*), we can express the following relation:

$$\mathbf{P} = [\rho_{ij}] = -\boldsymbol{\Delta}\boldsymbol{\Omega}\boldsymbol{\Delta} \tag{1}$$

where $\mathbf{P}$ denotes the partial correlation matrix and $\boldsymbol{\Delta} = diag(\frac{1}{\sqrt{\omega_{ii}}})$ [16]. Under the Gaussian assumption we can estimate efficiently the partial correlation matrix using the *glasso* model introduced in [10], that has the advantage of providing a sparse estimate of $\boldsymbol{\Omega}$ by penalizing the 1-norm of the precision matrix. The main disadvantage of this procedure is the normality assumption, therefore we use the alternative *tlasso* algorithm, that is based on *glasso* but relies on the assumption of multivariate t-Student distribution. *Tlasso* estimates are computed efficiently using an Expectation Maximization (EM) algorithm and in simulation studies proved to be more robust to misspecification and outliers in the data. We refer the reader to [10] and [6] for the technical details of *glasso* and *tlasso* respectively. Finally we underline that in this work we use the convenient representation of a network in terms of an *adjacency matrix* $\boldsymbol{A}$, i.e. a square matrix in which each entrance $[\boldsymbol{A}]_{ij} \quad \forall i \neq j$ represents the weight of the edges i, j and the elements on the main diagonal are equal to 0. In this work we use *tlasso* to estimate the network structure of the European banking system, assuming that partial correlations between financial time series (in particular CDS spreads) can represent a proxy of the the interconnections between banks and therefore the channel of propagation of financial distress as in [4].

2.2 Community Detection

A *community* in the field of complex network can be defined as a group of nodes that are more densely connected among themselves than with nodes outside the group. The problem of identifying the best community structure is well studied in the network literature (see for instance [7])[2]. We consider an optimization-based approach in which the optimal community structure is the one associated with the highest *modularity* [13], a quantity defined as follows. Given a partition $G = \{G_1, \ldots, G_p\}$ the modularity Q is:

$$Q = \frac{1}{2m} \sum_{i,j} \left(a_{ij} - \frac{s_i s_j}{2m} \right) \mathbb{I}_{[g_i = g_j]} \tag{2}$$

where a_{ij} is an element of the adjacency matrix $\boldsymbol{A}$, s_i is the strength of node i, $m = \frac{1}{2}\sum_{i,j} a_{i,j}$, g_i is the group in the partition in which the element i belongs and $\mathbb{I}_{[g_i = g_j]}$ is 1 if $g_i = g_j$ and 0 otherwise. Modularity can assume values between -1 and 1, with positive and high values denoting a good division of the network into communities. The procedure proposed in [13] identifies the optimal partition using a greedy optimization that, starting with each vertex being the unique member of a community, repeatedly joins together the two communities whose amalgamation produces the largest increase in modularity. This approach can be implemented efficiently on large networks and identifies automatically the optimal number of communities. Note that a positive value of modularity is not a sufficient condition for identifying a network divided in communities, therefore we need to test if it is the modularity is statistically significantly higher than the one of a random network. In particular we generate the random networks using a degree-preserving rewiring procedure [7].

3 Empirical Analysis

3.1 Data Description

Our dataset consists of 31 weekly time series of CDS (5 years maturity) of European financial institutions settled in 12 countries. They refer to CDS spreads quoted in Euro and they span the time period from January 2009 to June 2016. 20 of the banks in the sample belong to countries in the Eurozone, the other 11 are located in the United Kingdom, Sweden and Denmark. We observe that our database includes 85% of the banks with total assets over 500 billions that are under the European Central Bank (ECB) supervision and it is also consistent with the

[2]In the case of non-overlapping communities we can refer to the optimal community structure as *optimal partition*.

European Banking Authority (EBA) stress-test exercise 2016, representing 47% of the banks involved. For the analysis we consider the log-differences of CDS spreads, computing the partial correlation matrix from them using *tlasso* algorithm. We first estimate the network using the data of the entire sample period, and then we analyse the evolution over time using a rolling analysis using windows of 100 weekly observations each.

3.2 Empirical Results

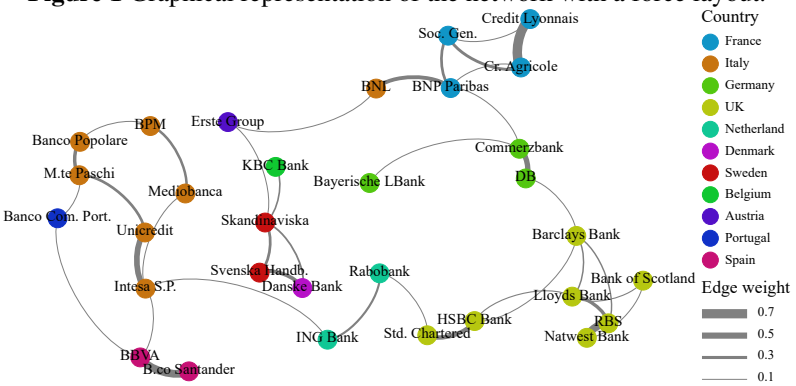
Static Analysis

Figure 1 shows the network represented using a force layout. The visual inspection denotes the clustering of banks in communities aligned with the national groups.

Table 1 shows the composition of the optimal communities identified by Newman's algorithm. The partition consists in five communities and it is possible to notice that it roughly overlaps with geographical divisions, confirming the results of the visual inspection of Figure 1. In particular, community 1 is composed uniquely by banks from Mediterranean Countries, community 2 by British and German banks, communities 3 and 4 include a more diversified group of banks from United Kingdom (UK), central and northern Europe and finally community 5 is composed by French banks (with the exception *BNL*, that is part of the French group *BNP Paribas* but is Italian).

Table 2 reports the value of modularity of the optimal community structure compared to two geographical partitions, one obtained grouping banks by country and one by grouping them in three broad geographical areas: Southern Europe, Central Europe and countries outside Eurozone. For comparison we also consider a partition based on the size of the banks, to check whether banks of similar dimension tend to connect to each other[3]. For each indicator we compute a confidence level based on the empirical distribution of the indicator computed on 1000 random rewirings of the network. We observe that the modularity of the optimal partition is equal to 0.461 and statistically significantly different from the null model with a confidence level higher than 99%, confirming that the network is characterized by a relevant division in communities. We also see that the modularity of geographical partitions (0.352 and 0.335 for the country partition and the area partition respectively), although smaller than the optimal one, are rather high and statistically significant, indicating that the geographical divisions represent a relevant feature of the banking network. Concerning the partition by size, although the modularity is positive and statistically significant, it has a much smaller value compared to the other partitions (0.087), suggesting that is a less relevant factor.

Figure 1 Graphical representation of the network with a force layout.



Dynamic Analysis

We perform a rolling analysis to monitor the evolution of the community structure over time. In particular, we consider the evolution of modularity as presented in Figure 2. We see that the modularity of the optimal partition is highest in the period corresponding to the Sovereign crisis, decreases from mid-2012 and then grows again in recent years. The pattern is similar for the geographical partitions, while modularity of the partition generated by the size show a moderate increase across the time period. The high level of modularity during the crisis is consistent with the sovereign-driven nature of the European crisis: the increased relevance of country risk leads to a decrease in confidence in the transnational interbank market, and thus to a "flight to safety"and a tightening of national banking systems. The rise in modularity in the last part of the sample may be related to the low level of the interbank interest rates in recent years, that makes less convenient for banks in core Countries to lend to banks in peripheral Countries, exacerbating the division among national banking systems. In an unreported test, we also

[3]In particular we defined 5 classes of homogeneous size based on the *total assets* based on 2015 balance sheet.

Community	Number of banks	Constituents
1	9	M.te Paschi(IT), Banco Popolare(IT), BPM(IT), Intesa S.P.(IT), Mediobanca(IT), Unicredit(IT), B.co Santander(ES), BBVA(ES), Banco Com. Port.(PT)
2	7	Commerzbank(DE), DB(DE), Barclays Bank(UK), Bank of Scotland(UK), Lloyds Bank(UK), Natwest Bank(UK), RBS(UK)
3	6	KBC Bank(BE), Bayerische LBank(DE), Erste Group(AUT), Skandinaviska(SE), Svenska Handb.(SE), Danske Bank(DK)
4	4	Rabobank(NL), ING Bank(NL), HSBC Bank(UK), Std. Chartered(UK)
5	5	BNL(IT), BNP Paribas(FR), Cr. Agricole(FR), Credit Lyonnais(FR), Soc. Gen.(FR)

Table 1 Constituents of Optimal Communities

Partition	Modularity
Optimal partition	0.461***
Countries	0.352***
Geographical area	0.335***
Total assets	0.087***

Table 2 Modularity of 4 partitions. ***,**,* refer to confidence level of 99%, 95% and 90% respectively.

measure the stability of the community structure by testing how well the optimal partition in a given estimation window can describe the community structure in a future window. We found that the optimal partition, despite the variations in the modularity over time, is characterized by a great stability. The results are available from the author upon request.

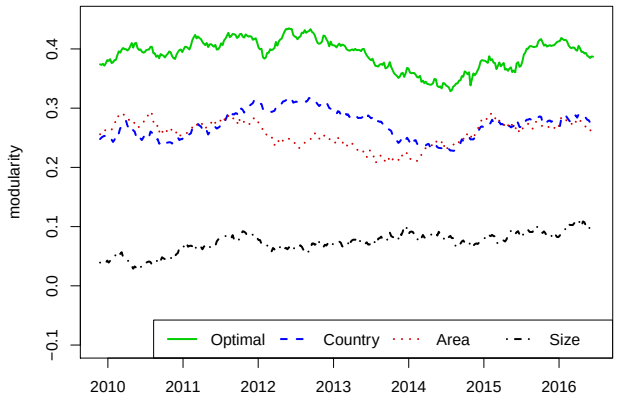


Figure 2 Evolution of Modularity over Time.

4 Economic Implications and Applications to Risk Management

It is well known that the presence of a community structure influences greatly the diffusion of epidemics and, in general, the behaviour of dynamic processes in complex network. Epidemic models find interesting applications in the economic literature, and have been used by several authors to model the diffusion of financial contagion [8]. However, to the best of our knowledge, the role of a community structure has not been explicitly studied in this literature. The results we present here describe the presence of a stable and significant community structure in the data and suggest to further study the topic, opening a new line of research focused in the study of the effect of community structure on the diffusion of financial distress.

The analysis of contagion dynamics are outside the scope of this work, however we propose here a simple procedure to use the information regarding the community structure for the assessment of the role of each bank

in the system. The procedure is based on the decomposition of *strength centrality*, an indicator computed for each node as the sum of the weights of the edges connected to it. We can think to a node characterized by a higher centrality as more interconnected, and therefore more systemically relevant. If the network is characterized by a community structure, a high centrality of a node could be associated to strong bonds to nodes in the same communities or to bonds to nodes in different ones. In a banking system this distinction is particularly relevant for the management of systemic risk: for instance a bank that has a high level of interconnectedness, but whose connections span mostly in a limited neighbourhood, would be less relevant in terms of systemic risk compared to a bank with broader interconnections, that could represent a "bridge" for financial contagion[4]. In particular we decompose *strength centrality* in two components: *strength inside* and *strength outside*.

$$Str_i^I = \sum_j a_{ij} \mathbb{I}_{[g_i = g_j]} \tag{3}$$

$$Str_i^O = \sum_j a_{ij} \left(1 - \mathbb{I}_{[g_i = g_j]}\right) \tag{4}$$

where a_{ij} is an element of the weighted adjacency matrix, G is the optimal partition of the network, g_i is the group in the partition in which the element i belongs and $\mathbb{I}_{[g_i=g_j]}$ is 1 if $g_i = g_j$ and 0 otherwise.

Figure 3 reports the decomposition of *strength centrality*. We can see that for most of the banks the *inside* component (dark) is particularly relevant, representing the largest part of the total *strength*, while the *outside* component (light) is in many cases marginal. Focusing on individual banks, we can use the decomposition to enrich the information coming from centrality measures. For instance, in the first community we notice that *M.te Paschi* and *Mediobanca* have similar value of *strength centrality*, the decomposition however shows that the former is mostly exposed to banks in the same community, while the latter has connection that span more internationally. Comparing to the state-of-art approach to financial regulation, we can make a relation between this indicator and the Global systemically important banks (G-SIB) assessment methodology proposed by Basel Committee. Two of the criteria for the identification of G-SIBs are related to the international exposure of a bank: cross-jurisdictional claims and cross-jurisdictional liabilities. The idea is that the greater a bank's global reach, the more difficult it is to coordinate its resolution and the more widespread the spillover effects from its failure. Our indicator is constructed on different basis, but provides similar information, showing the extent of the interconnection of the node in the network and the potential footprint of a credit event of an institution in the system using a network based model based on easily available data. Finally we underline that the *strength inside* and *strength outside* could find application in the early warning literature for financial distress or for the definition of network based capital requirements for banks (see [3]).

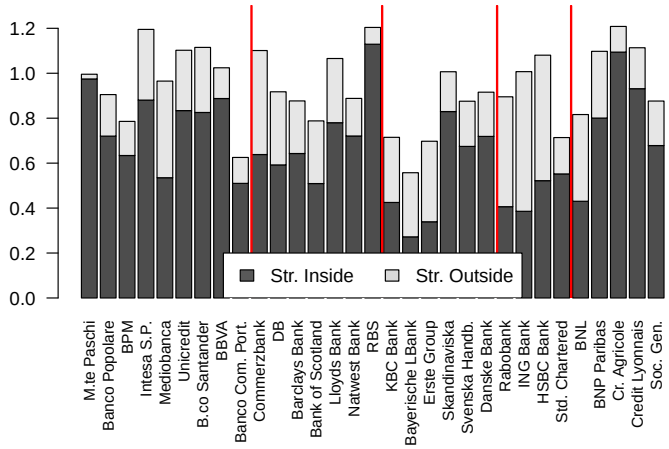


Figure 3 Decomposition of *strength centrality* in *strength inside* and *strength outside*. Vertical red bands divide the optimal communities.

[4]Note some centrality measures such as *eigenvector centrality* and *betweenness centrality* already allow to measure the relevance of each country in a more meaningful way, for instance weighting more the connections to more important nodes. Our approach is complementary to these measures, and we claim that it is more flexible, providing more insights on the role of each bank in the network.

5 Concluding Remarks

This work estimates the network structure of the European banking system on the basis of partial correlations using the *tlasso* algorithm. The analysis is focused on the identification of a community structure in the network and its influence on contagion risk. The results support the presence of a strong community structure and indicate that this structure is largely aligned to the geographical distribution of the banks. Furthermore, we propose a simple decomposition of *strength centrality* based on the community structure and we highlight its usefulness for assessing the role and importance of each bank in the network. The work opens new research questions regarding the role of community structure in the diffusion of financial contagion, that haven't been addressed yet in the literature.

Acknowledgements

I would like to thank Sandra Paterlini and Rosella Giacometti for introducing me to the topic and for the valuable help. All errors are mine.

References

[1] Aldasoro, I. and Alves, I., Multiplex interbank networks and systemic importance An application to European data, *Journal of Financial Stability*, forthcoming.

[2] Alter, A. and Schler, Y. S. , Credit spread interdependencies of European states and banks during the financial crisis, *Journal of Banking & Finance*, 36(12), 3444–3468, (2012).

[3] Alter, A., Craig, B. and Raupach, P., Centrality–based capital allocations, *International Journal of Central Banking*, 11(3), 329–376, (2014).

[4] Anufriev, M. and Panchenko, V., Connecting the dots: Econometric methods for uncovering networks with an application to the Australian financial institutions, *Journal of Banking & Finance* 61, S241–S255 (2015).

[5] Craig, B. and Saldías, M., Spatial Dependence and Data-Driven Networks of International Banks, IMF working paper WP/16/184, (2016).

[6] Finegold, M. and Drton, M., Robust graphical modeling of gene networks using classical and alternative t-distributions, *The Annals of Applied Statistics* 5(2A), 1057–1080, (2011).

[7] Fortunato, S., Community detection in graphs, *Physics reports*, 486(3), 75–174, (2010).

[8] Hurd, T. R., Contagion! Systemic Risk in Financial Networks, *Springer*, (2016),

[9] Fabozzi, F. J., Giacometti, R., Tsuchida, N., The ICA-based Factor Decomposition of the Eurozone Sovereign CDS Spreads, *Journal of International Money and Finance*, vol 65, 1–23, (2016).

[10] Friedman, J., Hastie, T. and Tibshirani, R., Sparse inverse covariance estimation with the graphical lasso, *Biostatistics* 9(3), 432–441, (2008).

[11] Lauritzen, S. L., Graphical models, *Clarendon Press* vol 17 (1996).

[12] Mistrulli, P. E., Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns, *Journal of Banking & Finance*, 35(5),1114–1127, (2011).

[13] Newman, M., Fast algorithm for detecting community structure in networks, *Physical review E* 69(6), (2004).

[14] Puliga, M., Caldarelli, G., Battiston, S., Credit Default Swaps networks and systemic risk, *Sci. Rep.* 4, 6822; DOI:10.1038/srep06822 (2014).

[15] Schwarcz, S., Systemic risk,*Duke Law School Legal Studies Paper No. 163* , *97(1)*, (2008).

[16] Stevens, G. V. G., On the inverse of the covariance matrix in portfolio analysis, *The Journal of Finance* 53(5), 1821–1827, (1998).

Nonparametric Kernel Regression and Its Real Data Application

Tomáš Ťoupal[1], František Vávra[2]

Abstract. This paper deals with the problem of nonparametric kernel estimation, particularly nonparametric kernel estimation of the regression functions. This nonparametric approach is useful in the case, when we need to find some relation between a pair of random variables for further analysis.
There are many fields of application in macroeconomics and therefore this paper is focused on estimates of the regression functions on some selected real data sets (number of deaths, marriages and births etc.) First, there is described nonparametric kernel estimation of the regression function with using Nadaraya–Watson approach and influences of the main parameters (smoothing parameter, kernel function etc.) on the properties of the regression function. Then, there is analyzed smoothing parameter and its estimation by different approaches (Penalty methods, RSS method, Cross-validation method and other proposed methods). The obtained results are applied and discussed on selected real data sets.

Keywords: Nonparametric kernel estimation, regression function, kernel function, smoothing parameter, Penalty methods, RSS method, Cross-validation method, MSE.

JEL Classification: C14, C51
AMS Classification: 62G08

1 Introduction

Nonparametric kernel estimates have become increasingly popular method within software development (not only) in the field of mathematical statistics. They can be used (not only) to estimate the unknown probability density functions or the cumulative distribution functions to obtain the main characteristics of random variables or to estimate the regression functions (regression curves) and then use them for example in the reliability estimation. One of the undisputed advantages of nonparametric kernel estimation is that an a priori information on the analyzed data sets is not required to obtain the estimation. The disadvantage in some cases may be higher time requirements depending on the parameters of the selected model.

The main goal of this paper is to analyze some of the used approaches to estimate the unknown smoothing parameter in nonparametric kernel regression and propose some robust approach. Then the obtained conclusions apply to the generated and real data sets including a summary of the recommendations and suitability of each method.

1.1 Nonparametric kernel estimation of probability density function

Let $\{x_i\}_{i=1}^n$ be an independent and identically distributed random sample of continuous random variable X of size n, then the approximation of the unknown density function, denoted as $\hat{f}(x;h)$, can be obtained as

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right), \qquad h > 0 \qquad n \in \mathbb{N}_+, \tag{1.1}$$

where h is a smoothing parameter called the bandwidth parameter. This parameter can be dependent on sample size n or on observations x_i and $k(x)$ is a kernel function. It is a continuous function satisfying the following conditions:

$$k(x) = k(-x),\ \int_{-\infty}^{+\infty} k(x)\,dx = 1,\ \int_{-\infty}^{+\infty} xk(x)\,dx = 0,\ \int_{-\infty}^{+\infty} x^2 k(x)\,dx = 1. \tag{1.2}$$

It means an even function that integrates to one with zero mean value and variance is one. Some often used and "common" kernel density functions analyzed here are shown in Table 1.

[1] NTIS – New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, Plzeň, Czech Republic, e-mail: ttoupal@kma.zcu.cz.
[2] Department of Mathematics, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, Plzeň, Czech Republic, e-mail: vavra@kma.zcu.cz.

Kernel functions	Description $k(x)$	Assumptions
Parzen (uniform)	$\frac{1}{2a}$	$-a \le x \le a;\ a > 0;\ a \in \mathbb{R}$
Epanechnikov	$b\left(1-\left(\frac{x}{a}\right)^2\right)$	$-a \le x \le a;\ a,b > 0;$ $a,b \in \mathbb{R}$
Triangular	$b\left(1-\frac{\lvert x\rvert}{a}\right)$	$-a \le x \le a;\ a,b > 0;$ $a,b \in \mathbb{R}$
Gauss	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$x \in \mathbb{R}$

Table 1 Used kernel functions

These kernel functions are only a few of all possible kernel functions and are suitable for the intended nonparametric kernel estimates because of their "simplicity".

1.2 Nonparametric kernel regression

The aim of this paper is to analyse the problem of regression with a focus on the use of nonparametric kernel estimation, i.e. a nonparametric approach is used against the parametric because of the flexibility of the method. This regression curve generally describes the relationship between a random variable (Y) and one or more non-random variables X.

There are n ordered pairs of observations $\{(x_i, y_i)\}_{i=1}^n$ of two variables X, Y. The aim is to approximate (describe) the relationship between both variables by using some function m. This relationship can be described as

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1, \dots, n, \tag{1.3}$$

where $x_i, y_i \in \mathbb{R}$ are $i-$th observations of both variables, $m(x)$ is the unknown, estimated and predicted regression function and $\varepsilon_i \in \mathbb{R}$ represents the $i-$th observation of unknown random variable ε satisfying the requirements:

$$\begin{aligned} E(\varepsilon_i) = 0, \quad D(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2, \quad i = 1, 2, \dots, n, \\ C(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0, \quad i, j = 1, 2, \dots, n; i \ne j. \end{aligned} \tag{1.4}$$

The regression curve estimation can be done by more approaches, where there are models based on the work of Priestley and Chaa (1972), Nadaraya and Watson (1964), Gasser and Müller (1979) and Cheng and Lin (1981). These estimates are characterized by similar properties at equidistant observation points. Because of the frequent use in practice, Nadaraya-Watson's estimate is analysed here, for example see [1].

Let X and Y are i.i.d. random continuous variables for n observations with two–dimensonal joint density function $f(x, y)$, where $x, y \in \mathbb{R}$, Y is the dependent variable and X is the independent variable. Then there is a function in the sense $\mathrm{E}(Y|X = x) = m(x)$ and the Nadaraya-Watson regression model is

$$\hat{m}(x; h) = \begin{cases} \dfrac{\frac{1}{n}\sum_{i=1}^n k\left(\frac{x - x_i}{h}\right) y_i}{\frac{1}{n}\sum_{i=1}^n k\left(\frac{x - x_i}{h}\right)}, & \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right) \ne 0, h > 0, \\ 0, & \text{others}, \end{cases} \tag{1.5}$$

where $\hat{m}(x; h)$ is the nonparametric kernel estimate of the regression curve $m(x), h > 0$ is the smoothing parameter, $n \in \mathbb{N}_+$ denotes the number of observations, $k(x)$ kernel function satisfying the above conditions and x_i, y_i are realization of X, Y.

1.3 Modelling and smoothing parameter estimation

The smoothing parameter affects (as well as the kernel density estimation) the "degree" of smoothing of the estimated regression function. The main problem in nonparametric kernel estimation of the density or distribution function and the regression function is the choice of the smoothing parameter h $(h > 0)$, where it is necessary to find a compromise between the "undersmoothed" and "oversmoothed" estimation, see for example [2]. The lower values of this parameter indicate a higher influence of the observations around the estimated value, and conversely, the higher values indicate a lower influence of the observations around the estimated value. It is also important here the knowledge of the theoretical model for possible estimation. Due to this fact, there are analyzed only the estimates of the unknown smoothing parameter without a priori information.

The following acceptable approaches are based on the Average Mean Squared Error (AMSE), especially on the asymptotic equivalent Average Squared Error (ASE)

$$ASE(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{m}(x_i; h) - m(x_i)\right)^2, \quad i = 1, \dots, n, \tag{1.6}$$

which depends on the unknown value of the predicted regression function. This deficiency can be replaced by y_i for obtaining the Residual Sum of Squares (RSS) minimizing the optimal width of h

$$RSS(h) = \frac{1}{n}\sum_{i=1}^{n}(\hat{m}(x_i; h) - y_i)^2, \quad i = 1, \dots, n. \tag{1.7}$$

The known method for smoothing parameter estimation is the Cross-Validation method. This method uses estimates without the gradual use of each observation and is based on the Residual Sum of Squares [3], i.e.

$$\hat{h} = \arg\min_{h \in H_n} CV(h), \tag{1.8}$$

where H_n is the set of all values of the smoothing parameter and $CV(h)$ is the Cross-Validation function

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}[\hat{m}_{-i}(x_i; h) - y_i]^2. \tag{1.9}$$

Here $\hat{m}_{-i}(x_i; h)$ is Nadaraya-Watson's kernel estimate of the regression function at point x_i without the same observation and the other variables are the same as before. Another possibility to find the optimal value of the smoothing parameter h is to use the Penalization methods. The main principle is based on the multiple of the RSS function (1.7) by a "suitable" function, which for small values h obtains large values and converges to 0 for a large values h. This function is called penalizing function because it penalizes small and large values of the smoothing parameter h. Possible penalizing functions are listed in the following table [4].

Name	Penalization function
Generalized cross-validation	$\varphi(z) = \frac{1}{(1-z)^2}$
Akaike's Information Criterion	$\varphi(z) = e^{2z}$
Finite Prediction Error	$\varphi(z) = \frac{1+z}{1-z}$
Shibata's model selector	$\varphi(z) = 1 + 2u$
Rice's bandwidth selector	$\varphi(z) = \frac{1}{1-2z}$
ET bandwidth selector	$\varphi(z) = e^{\frac{4}{\pi}tg\frac{\pi}{2}z}$

Table 2 Possible penalization functions

Another possibility to find the optimal value of the smoothing parameter h is to use a special optimal function with regularization parameter [6]. The main principle of this method is to make the kernel estimate as close as possible to obtained observations (to minimize RSS) but the resulting curve must be smoothed. For the calculation is chosen one where both criteria have the same weight and the main formula is

$$\hat{h} = \operatorname*{argmin}_{m}[RSS + f(m(x))], \tag{1.10}$$

where the smoothness is obtained by the sum of absolute values of the first or second differences and both parts of the formula are divided by their norms for the same weight. These approaches are widely used, but as it can be seen from the following figures, they are in some sense an "unsmoothed" estimate. For this reason, the main aim is to propose an approach for the smoothing parameter estimation with a "smoother" estimate of the regression curve. One possible approach is based on the variance equality. It is obtained on the base of nonparametric kernel density estimation. Here is required the same estimate of the variance as the sample variance and

$$\hat{h} = \sqrt{\frac{s_n^2}{n}}, \quad n \in \mathbb{N}_+, \tag{1.11}$$

where s is the sample standard deviation. Another possible approach to estimation the smoothing parameter is given by the following formula

$$\hat{h} = sn^{\frac{-2}{5}}, \quad n \in \mathbb{N}_+. \tag{1.12}$$

This approach is based on the probability convergence condition [5]

$$\lim_{n\to\infty} P\{sup_x|f(x) - \hat{f}(x)| \geq \varepsilon\} = 0. \quad (1.13)$$

It is important to note here, that when the above approaches are used to estimate the smoothing parameter, it is assumed, that the unknown smoothing parameter h is determined by a real number (this is a deterministic value), but in this case the value of this parameter has the character of a random variable, which is a general problem.

2 Data sets application

2.1 Generated data set

First, there is one generated function used to presentation of the obtained results, more precisely it is a function

$$y = \sin(2x) + \varepsilon, \quad \varepsilon \sim N(0,1). \quad (1.14)$$

The following figure shows this function and nonparametric kernel estimates of the regression curve using the above-mentioned kernel functions for the same (not optimized) smoothing parameter $h = 0{,}4$.

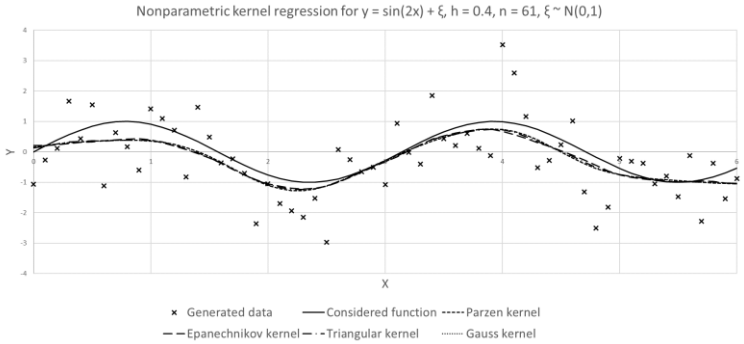


Figure 1 Generated data set with different kernel functions

Here it is possible to notice, that the used kernel function does not have a significant effect on the shape of the regression curve. The figure below illustrates the effect of the smoothing parameter on the regression curve approximation. Here are shown nonparametric kernel estimates of the regression curves using the Gaussian kernel function for different values of the smoothing parameter h. The effect of this parameter is noticeable, i.e. if it is low, then the estimate is "undersmoothed" and the resulting regression curve copies the obtained observations, and vice versa, if it is high, then the estimate is "oversmoothed" and the estimated regression curve does not reflect possible influences.

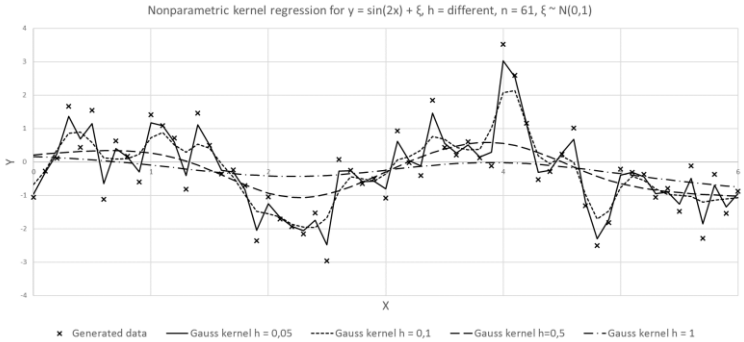


Figure 2 Generated data set with different smoothing parameters

2.2 Real data sets

The proposed methods are now applied to selected real data sets, more precisely to the data sets with the number of marriages, births and deaths for further analysis. These data are obtained in the period from 1785 to 2015 and are plotted in the following figure.

Figure 3 Selected real data sets

Comparison of different approaches to estimating the smoothing parameter are shown for a data set of marriages in CR.

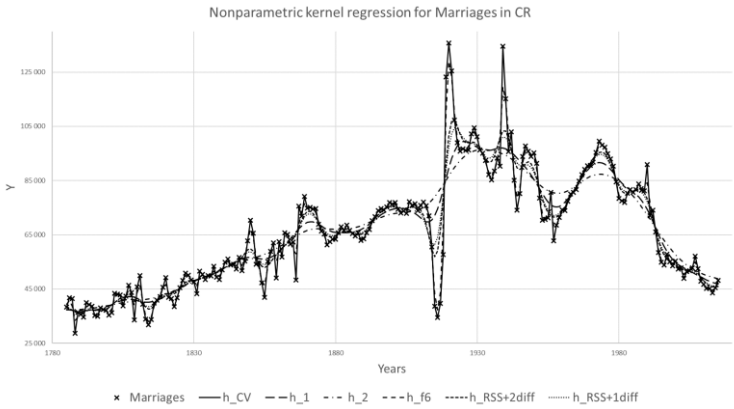


Figure 4 Nonparametric kernel regression for marriages using different approaches

Here it is possible to see the differences in approaches used to estimate the smoothing parameter. The most smoothed results are obtained by formulas (1.11) and (1.12). These approaches are derived from the nonparametric kernel density estimation and reach the maximum values of MSE. The least smoothed results are obtained by CV(h) function and penalisation functions (only one with the highest value is shown in the figure for clarity). These approaches reach the lowest values of MSE. The compromise between these two approaches is the estimation of the smoothing parameter according to formula (1.10). For further analysis can be required the shapes of obtained regression curves (e.g. to estimate some mathematical model for prediction). For this reason, two selected data sets are the number of births and deaths (for their interesting interpretation). The obtained regression curves with minimum values of MSE (CV method), including the difference (Births – Deaths), are shown in the figure below.

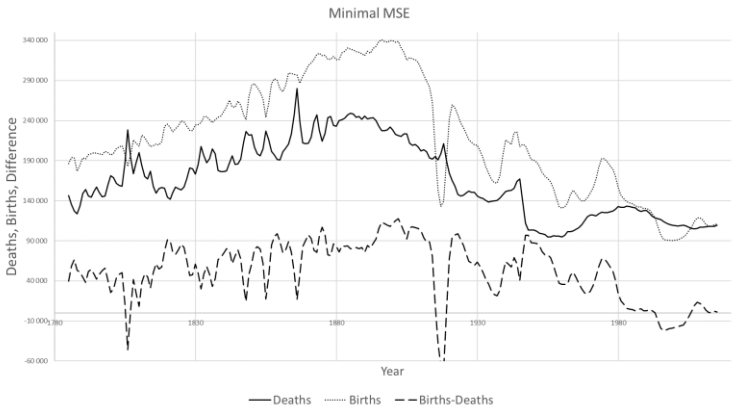


Figure 5 Estimated regression curves with minimum MSE

On the other hand, the regression curves with maximum values of "smoothing" are shown in the following figure (i.e. it means minimum values of the differences) and are obtained by the formula (1.12).

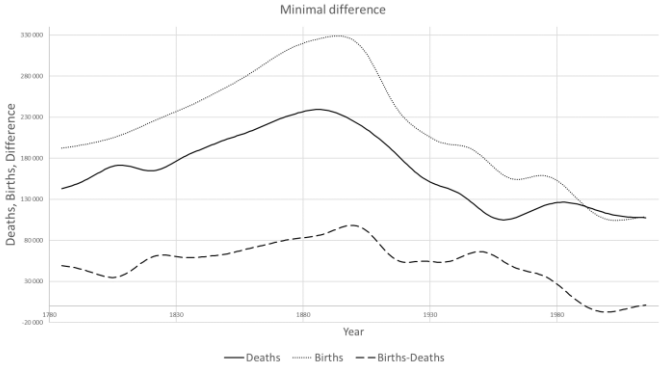


Figure 6 Estimated regression curves with minimum values of the differences

This approach is very useful for the following mathematical or statistical description, for example possible estimation of trend components. Finally, there are plotted regression curves, where the above quality parameters are minimalized (MSE and differences). This is a certain compromise between "smoothing" and "tightness".

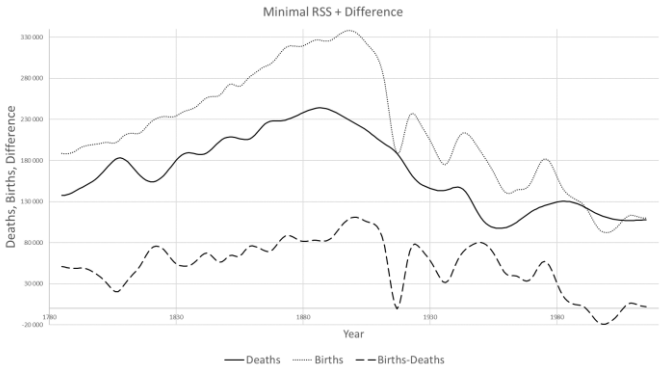


Figure 7 Estimated regression curves with minimum MSE and differences

3 Conclusion

This paper describes possible (not only) approaches to estimating the smoothing parameter. There are approaches used in practice (CV method, RSS, Penalty methods) and proposed new possible approaches based on the derivation of density function. At the same time, there is one approach minimizing "smoothing" and "tightness. The choice of some method to estimating the smoothing parameter is generally a subjective matter and each approach gives different results on different data sets.

Acknowledgements

This publication was supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports.

References

[1] Bierens, H. J.: Kernel Estimators of Regression Functions, kap. 3. In: *Advances in Econometrics: Fifth World Congress*, sv. I. BEWLEY, T. F. ed. New York: Cambridge University Press, 1987, p. 99-144. ISBN 0521345529.

[2] Ťoupal, T.: *Neparametrický odhad spolehlivosti a odhad trendové složky*. Plzeň, 2013. Disertační práce. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Katedra matematiky.

[3] Hardle, W.: *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1990.

[4] Koláček, J.: *Jádrové odhady regresní funkce*. Brno, 2004. Disertační práce. Masarykova univerzita, Přírodovědecká fakulta.

[5] Devroye, L., Gyorfi, L.: *Nonparametric Density Estimation: The L1 View*, John Wiley, New York, 1985.

[6] Šedivá, B., Marek, P.: *Term structure of interest rates: comparison of the Czech Republic and Germany*. In 33rd International Conference Mathematical Methods in Economics, MME 2015, Conference Proceedings, Cheb, University of West Bohemia, 2015. p. 787-792. ISBN: 978-80-261-0539-8.

Concept of Income Inequality Gap

Kamila Turečková [1], Eva Kotlánová[2]

Abstract. This article analyses through concept of income inequality gap impact of economic policy on reducing income inequality in society. Income inequality gap presents a difference between levels of income inequality measured on gross money income data and net money income data. This gap is determined by specific areas of social policy, concretely this gap considers on a system of mandatory contributions and income taxes. Via this income inequality gap we can measure success of redistribution function of economic policy that means the success of social policy in reducing income inequalities in society. If the difference between levels of both income inequalities is low we can assume about not very effective redistribution function of economic-social policy and vice versa. Mathematically is the income inequality and income inequality gap expressed and measured by the method of non-weighted average absolute deviation. The article would be rather theoretical focused but is supplemented by practical example of analysis of income inequality gap based on data for on data for Czech Republic from years 2005-2015.

Keywords: Concept of Income Inequality Gap, Income Inequality, Economic-Social Policy, Method of non-weighted Average Absolute Deviation.

JEL Classification: C13, D31, I32, O15.

AMS Classification: 62P20

1 Introduction

Income inequality has become one of an important issue in countries all over the world. This term is closely associated with poverty, affects society negatively and it could create a space for social and economic discrimination [18]. Because of that and the consequences resulting therefrom, one of the tasks of economic policy especially via social policy is to reduce income inequality. Economic policy could influence gross money incomes through minimum wage and social incomes (not only in context of subsistence minimum) or it could influence difference between gross money incomes and net money incomes through mandatory contributions and income taxes [19].

The aim of this article is to introduce a concept of income inequality gap that could be a good instrument for analyzing effects of economic-social policy on reducing income inequality in society in particular economy. This gap is calculated as difference between values of indexes expressing level of income inequality measured from gross money income and net money income of households on, in our case, deciles level in selected period of time.

The level or value of income inequality is calculated by relatively simple method for measuring, expressing and analyzing income inequality which is denoted Method of non-weighted average absolute deviation here. For more information about this new method and its legitimacy see for example [8, 17, 18, 20 or 21]. This alternative method can expand the existing portfolio of methods for measuring and expressing income inequality between households in society next to well-known methods how to measure income inequality. Among them traditionally belongs one of the best known and used measures of income inequality Gini coefficient and its graphical representation – Lorenz curve [20], or other as Coefficient of income inequality S80/S20 (or Quintile share ration or S80/S20 Ratio), Atkinson index, Theil index, Robin Hood index and Variation coefficient [3, 5, 10, 11, 12, 16 or 22].

Social policy in the Czech Republic has a redistributive character. Social policy is mainly based on the principle of solidarity and has protection, redistribution and prevention functions [6]. The aim of social policy is to positively influence people's living conditions [15] Krebs (2015) argues that social policy is perceived as a set of activities, tools and measures which purpose is the response to unfavorable social conditions such as old age, illness, disability or unemployment and poverty [9]. In practice, it means also a system of social benefits and

[1]Silesian University in Opava, School of Business Administration in Karvina, Department of Economics and Public Administration, Univerzitní nám. 1934/3, Karviná 733 40, tureckova@opf.slu.cz.
[2]Silesian University in Opava, School of Business Administration in Karvina, Department of Economics and Public Administration, Univerzitní nám. 1934/3, Karviná 733 40, kotlanova@opf.slu.cz.

social incomes and system of health and social insurance. The goal of social policy is to create dignified conditions of life for all people in the economy (in the state). As Arnoldová (2012, p. 41) says: "... every individual has from his/her birth the right to receive income that is necessary for satisfying his/her basic life needs. It is provided from birth, regardless if he/she may or may not work and regardless of his/her financial situation. Its level corresponds to the definition of poverty, which the society receives" [2]. Here is meant minimum income what is claimable benefit, which guarantees each social state. In the Czech Republic there are two basic categories: subsistence guaranteed basic income and the minimum wage [2].

From this point of view the income inequality gap are dependent (in Czech Republic) on social insurance system (health insurance and social insurance), Tax bonus to child and Income tax of individuals (Natural persons) that is direct tax and is set by Act No. 586/1992 Coll. on Income Taxes, as subsequently amended. The area of income taxes provides the widest scope for influencing income inequality in society. Social insurance has two parts: pension and sickness insurance that are based on the insurance principle, this means that it is a regular payment of the contribution – insurance premium. From sickness insurance is provides 4 types of state cash benefits: sickness benefit, maternity benefit, attendance allowance, compensatory benefits in pregnancy and maternity [7].

The other area in which could be income inequality gap influenced is the state social support includes state social support benefits as child allowance, parental allowance, housing allowance, maternity grant and death (funeral) grant. State through social policy and via payment of these benefits takes over the responsibility for unfavourable social situation in families and it citizens [14]. Social incomes have relatively very rapid growth and there can be recognize increasing share of them on total cash income of Czech households. According to Nečadová (2011, p. 64) social incomes in the Czech Republic represent about 20% of all households' revenue [13]. The largest proportion of all social incomes has pension benefits - 73% of all social income, followed by state social benefits - 13%, health insurance benefits for about 8%, the benefits of social assistance and unemployment allowance for about 3%. Social support benefits form part of the social income so this is a way how could economic national authorities influence a gross money income. A specific social income is pension that is paid from the pension insurance in the Czech Republic. Pension insurance is compulsory for all economically active persons over 18 years. From this basic pension insurance is paid (old age) pension in all variants, invalidity, widow's and widower's and orphan's pension. Through these contributions to mandatory social security schemes it is possible (for national economic authorities) to influence a difference between gross and net money income and affect income inequality gap.

The text after this Introduction is organized subsequently: the second part introduces a sense of "term" income inequality gap and explains the theoretical approach of concept of income inequality gap based on method of non-weighted average absolute deviation. There is highlighted role of particular tools of economic-social policy that could contribute to reducing income inequality and widen income inequality gap in economy as well too. The next section contains the practical example of empirical analysis and results of income inequality in context of functioning of economic-social policy and there is also given information about used data and methodology. The major findings are summarized in the conclusion.

2 Theoretical approach to concept of income inequality gap

Income inequality is a natural part of every human society and could reflect a measure of poverty and redistribution of income. There is used the Method of non-weighted average absolute deviation (shortly method of average deviation) for expressing level of income inequality in the society. This method reflects the degree of variability, defined as the arithmetic mean of the absolute deviations of individual values of observed indicators from the ideal value. This value chosen here understands the value for the ideal distribution of income in society, ie. the value of expressing absolute equality in income for each inhabitant [21]. Own value of non-weighted average absolute deviation that presents a value of income inequality (d_{II}) we obtained from the formula 1:

$$d_{II} = \frac{\sum_{i=1}^{p} |x_i - x_{io}|}{N_i} \tag{1}$$

where: $d_{II} \in (0;100)$; x_i presents the *i-th* indicator (particular variable); x_{io} is the ideal value; N_i presents the number of values of *i-th* indicator that we have available.

If there is perfect income equality in the society the value of income inequality index (d_{II}) come out zero. The higher value of index means a bigger difference of income between the richest and poorest households in society what is unwanted for national economic authorities and people too.

Income inequality gap is author's term to express difference between index values of income inequality counting from Gross money income (d_{II}^{GMI}) and Net money income (d_{II}^{NMI}). Income inequality gap (GAP_{II}) can be calculated by the simple equation (2):

$$GAP_{II} = d_{II}^{GMI} - d_{II}^{NMI} \tag{2}$$

Income inequality gap is a value that expresses "activities" of economic policy which are used by national economic authorities to mitigate income inequality in the society. If this policy leads to a reduction of the gap we can talk about the income pro-equal social policy, conversely we talk about income dis-equal social policy because the redistribution function of economic policy is not set adequately to reduce income inequality between all social groups of people in the particular economy and society. The theoretical definition of the concept of income inequality gap is displayed through a schema without values in Figure 1.

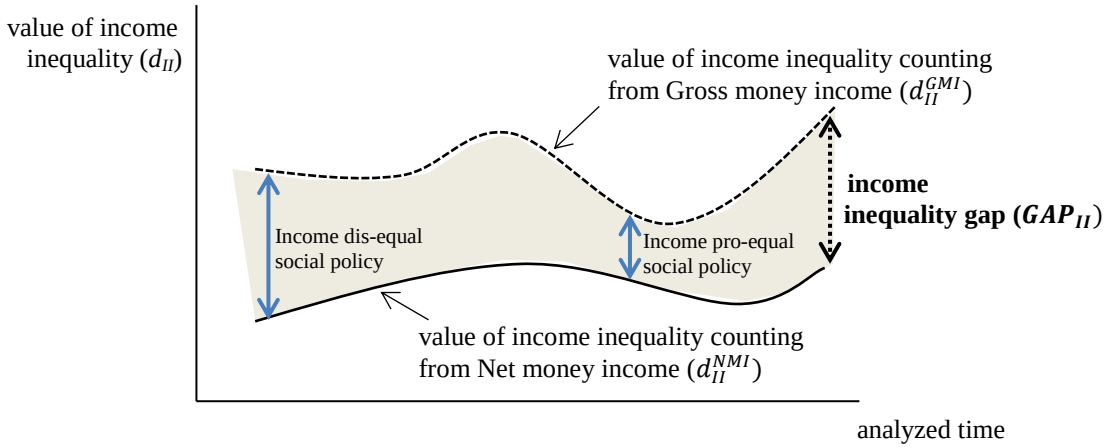


Figure 1 Income inequality gap in context of economic-social policy

Income inequality gap takes values from zero to the higher limit defined as value of the variable d_{II}^{GMI}; $GAP_{II} \in \langle 0; d_{II}^{GMI})$ when $GAP_{II} = 0$ means that there is no difference between gross and net money income and there are not any mandatory contributions or income taxations. On the other side, when GAP_{II} is as large as variable d_{II}^{GMI} it indicates that national authorities through its instruments are trying to reduce income inequality and the distribution of net money incomes between households is steady – uniform. From this point of view more active economic (social) policy (pro-equal social policy) means higher value of income inequality gap or increase of the gap over time. Providing social income (state social support benefits and other social subsidy) can be projected directly into decline of the index value expressing income inequality [19].

National authorities could reduce income inequality in many ways. In our point of view it could be done via: (I) providing social income, (II) determining contributions to mandatory social security schemes and (III) fixed income tax.

Social income and set basic income or the minimum wage (minimum income) by an act exerts influence over gross money income when gross money income depends on salary and private financial benefit and has to be higher as minimum income. On the other hand net money income depends on gross money income, of course, and also on reducible personal allowances, tax bonus to child, contribution to mandatory social security schemes and income tax.

Social policies with regard to the foregoing affect the level of income inequality through social incomes and minimum income while the income inequality gap depends on structure of contributions to mandatory social security schemes and establish system of income taxes. Reducible personal allowances (bonus) cover all income groups of people therefore it not interested us in context of impact of social policy on income inequality. In a simplified way the income inequality gap depends on system of contribution to mandatory social security schemes and income taxes and on tax bonus to child.

Relevant and "required" impact of economic policy via system of contributions to mandatory social security schemes and income taxation can be "measured" through income inequality gap. If the income inequality gap has low value, it means that the difference between gross and net money income is small and economic policy is in this area not considerable (we can used term income pro-equal social policy) and vice versa (so we talk about income dis-equal social policy). If we are assuming that the goal of economic policy is to reduce income inequality, it is desirable to minimalize this income inequality gap and levels of both income inequality (work on gross money income and net money income) indexes are very similar to each other.

3 Empirical analysis of income inequality gap: case study for Czech Republic in years 2005-2015

From a methodological perspective, the work is based on secondary data gained by Czech Statistical Office [4], concretely from the Catalogue of Products: Household Income and Living Conditions; Households total by net money income per person – deciles; Household composition and per capita annual income (%, CZK). This statistic provides data on decile scale and summary of all categories of income in Czech Republic. For purpose of calculating income inequality gap is especially important data about Gross money income, Social income, Contributions to mandatory social security schemes, Income tax and Net money income. The covered period includes years 2005-2015 because of missing credible data which is not available for a longer period[3]. Calculations of value of income inequality and value of income inequality gap are based on calculations using formulas (1) and (2). These types of measurement were described in the text above. The software used was MS Excel. All calculations and graphical analysis is author´s own.

The development of amounts of value of income inequality counting from Gross money income and value of income inequality counting from Net money income and income inequality gap is shown below in Table 1.

Index/Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
d_{II}^{GMI}	16.46	16.20	16.14	15.75	15.73	15.55	15.67	15.48	15.39	15.70	15.50
d_{II}^{NMI}	14.68	14.45	14.24	13.95	14.06	13.97	14.06	13.87	13.76	14.04	13.79
GAP_{II}	1.79	1.75	1.90	1.80	1.66	1.58	1.61	1.61	1.63	1.66	1.72

Table 1 Income inequality indexes and Income inequality gap, 2005-2015.

As is shown in Table 1 income equality gets better improves (by about 1 index point) while income inequality gap did not considerably change during analyzed time. In context of theory there were no significant impacts to influence income inequality on net money income level from the side of national-economic authorities. The graphical development of all three indicators is shown in Figure 2.[4]

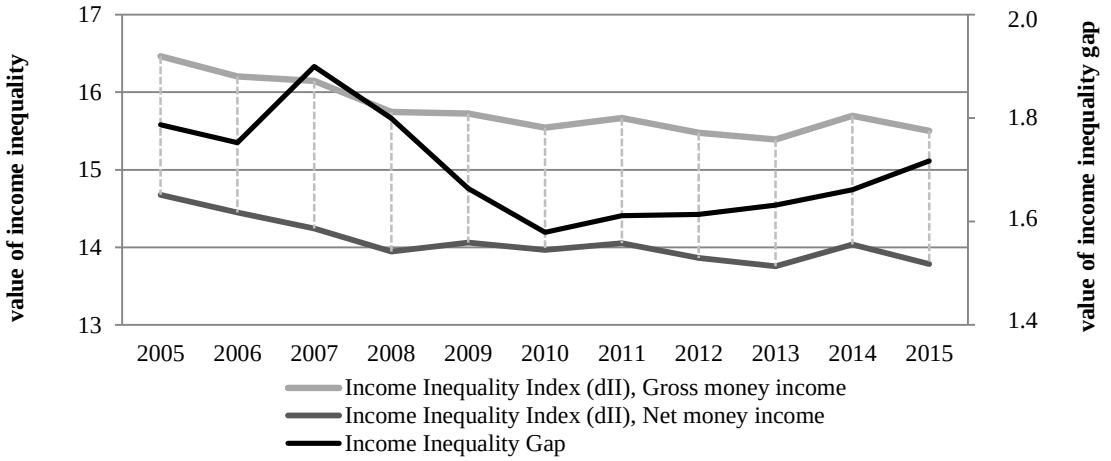


Figure 2 Development of Income inequality indexes and Income inequality gap, 2005-2015.

4 Conclusion

The aim of this short paper was to introduce the concept of income inequality gap what could be a good instrument for success expression of economic-social policy in reducing income inequality in society. So this income inequality gap is affected by tools of economic, especially social policy, concretely via income taxation of individual person, contribution to mandatory social security schemes and tax bonus to child. Although is this article and content of income inequality gap do for Czech region it may be after relevant adjustments overcome

[3] It is necessary to emphasize that the Statistics on Income and Living Conditions are published for one year, for example for 2006, but in statistical tables are presented data for year before, i.e. for 2005. Our paper is written for period of year 2005-2015 but empirical analysis is done for years 2004-2014.

[4] Present analysis is shown here only as an example. For completed and more detailed analysis of impact social policy on income inequality see [19].

to other countries. The article was supplemented by example of income inequality gap that was done for Czech Republic for and covered period was 2005-2015. It was calculated that this gap acquire values from 1.58 (2010) to 1.9 (2007). The calculation of income inequality gap is based on difference between values of income inequality index based on Gross money income and Net money income. These values were counted via method of non-weighted average absolute deviation. Data was gained by Czech Statistical Office.

The results of the presented analysis are the groundwork for a further theoretical and practical research of the role of social policy in affecting income inequality in society.

Acknowledgements

This paper was supported by the project SGS/13/2015 "Influence of Selected Macroeconomic and Microeconomic Determinants on the Competitiveness of Regions and Firms in Countries of the Visegrad Group Plus".

References

[1] Act No. 586/1992 Coll. on Income Taxes, as subsequently amended [online]. [2016-06-22]. Available from: http://www.czechlegislation.com/en/586-1992-sb.
[2] Arnoldová, A.: *Sociální zabezpečení I*. Praha: Grada. 2012.
[3] Atkinson, A. B.: On the Measurement of Inequality. *Journal of Economic Theory* **2** (1970), 244 – 263.
[4] ČSÚ: Czech statistical office: Household Income and Living Conditions 2005-2015. [online]. [2016-06-10]. Available from: https://www.czso.cz.
[5] Dalton, H.: *The measurement of the inequality of incomes*. [online]. [2014-08-19]. Available from: http://www.ophi.org.uk/wp-content/uploads/Dalton-1920.pdf.
[6] Duková, I., M. Duka and I. Kohoutová: *Sociální politika*. Praha: Grada. 2013.
[7] Hejkal, T.: *Nemocenské, důchodové a zdravotní pojištění s komentářem a příklady*. Praha: Polygon. 2004.
[8] Kotlánová, E.: Income Inequality and Economic Growth. In: *Proceedings of the 13th International Conference Economic Policy in the European Union Member Countries* (Machová, Z., and Tichá, M., eds.). VŠB-TU EKF Ostrava, Ostrava, 2015, 288-294.
[9] Krebs, V.: *Sociální politika*. Společnost Wolters Kluwers (ASPI). 2015.
[10] Kuznets, I.: Economic Growth and Income Inequality. *American Economic Review* **65** (1), 1955, 1-28.
[11] Lapáček, M.: *Příjmová nerovnost a ekvivalenční stupnice*. 2007.[online]. [2014-08-10]. Available from: http://nf.vse.cz/download/veda/workshops/inequality.pdf.
[12] Litchfield, J. A.: Inequality: Methods and Tools. Text for World Bank's Web Site on Inequality, Poverty, and Socio-economic Performance. [online]. [2014-08-22]. Available from: http://www.worldbank.org/poverty/inequal/index.htm.
[13] Nečadová, V.: *Sociální politika - 2. část – aplikace*. Jihlava: Vysoká škola polytechnická. 2011.
[14] Paulusová, M. a A. Vašíčková: Sociální pomoc ze strany státu neúplným rodinám s dětmi. *Sociální práce/sociálna práca 2/2014* (2014), 40-43.
[15] Potůček, M.: *Veřejná politika*. Praha: Sociologické nakladatelství. 2005.
[16] Schutz, R.:On the Measurement of Income Inequality. *The American Economic Review* **41** (1), 1951, 107 – 122.
[17] Turečková, K. and E. Kotlánová: Method of Non-weighted Average Absolute Deviation in Context of Measuring Income Inequality by Gini Coeffi-cient: case study of Visegrad group countries. In: *Conference Proceedings: 33rd International Conference Mathematical Methods 2015*. Plzeň: University of West Bohemia. 2015. 835-840.
[18] Kotlánová, E. and K. Turečková: Income Inequality in V4+ Countries at NUTS 2 Level. In: *Conference Proceedings from 34nd International Conference Mathematical Methods in Economics MME 2016*. Liberec: TU Liberec, 2016. 442-447.
[19] Turečková, K. and I. Buryová: Economic policy in context of income inequality gap: case study of Czech Republic in 2005-2015. In*: Proceedings of 14th International Scientific Conference Economic Policy in the European Union Member Countries. Karviná*: SU OPF Karviná, 2016. 777-785.
[20] Turečková, K., 2015. Income Inequality by Method of Non-weighted Average Absolute Deviation: Case Study of Central and Eastern European Countries. *Equilibrium* **10**(4), 2015. 99-110.
[21] Turečková, K.: Measurement of income inequality by method of non-weighted average absolute deviation: Case of South European Countries. In: *Proceedings of the 12th International Conference Liberec Economic Forum 2015*. Liberec: TU Liberec, 2015. 438-446.
[22] Wolff, E. N.: *Poverty and income distribution*. 2nd Ed. Chichester, UK: Wiley-Blackwell, 2009.

On Modelling of the Development of Turnover in Services in Slovak Republic: Tourism Approach

Petra Vašaničová[1], Eva Litavcová[2], Sylvia Jenčová[3]

Abstract. Tourism plays an important role in the Slovak Republic because of its economic potential and the employment potential as well as its social and environmental impact. It means that there is the potential to contribute to employment and economic growth, as well as the development of rural, peripheral and less developed areas. There are many indicators or variables using which one can models the development of tourism. Widely used are number of arrivals or nights spent at tourist accommodation establishments. To explore tourism can be also used other forms of indicators. One of them is turnover in services which is divided by classification of economic activities (NACE Rev. 2) and obtains from Structural business statistics. The aim of this paper is to model the development of turnover in services of economic activities air transport, accommodation, travel agency, tour operator reservation service and related activities in Slovak Republic. For that purpose we use exponential smoothing and Box - Jenkins methodology and monthly data from January 2000 to January 2017.

Keywords: tourism, turnover in services, ARIMA models, exponential smoothing.

JEL Classification: C22
AMS Classification: 91B84

1 Introduction

Tourism is one of the most developed sectors in the world's economy. The number of tourists is steadily increasing, thanks to which the number of employees in this sector is growing, as well as the total revenues from tourism. Even in Slovakia, tourism plays an important role, because it is a key sector in many regions of Slovakia [10].

The aim of this paper is to model the development of turnover in services of economic activities such as air transport, accommodation, travel agency, tour operator reservation service and related activities in Slovakia. This may be useful contribution to existing literature connected with modelling of tourism development, in which as indicators which explore tourism are widely used the number of arrivals or the nights spent at tourist accommodation establishments. For modelling and forecasting tourism development it this study we use two models: the exponential smoothing model, and the Box - Jenkins Seasonal Autoregressive Integrated Moving Average (SARIMA) model.

Mentioned methodology was used in number of tourism studies. For example, Lee, Song, Mjelde [7] predicted the number of visitors in Korea in the context of an international tourism Expo by using exponential smoothing and ARIMA model. ARIMA model is used in [5] to forecast overnight stays and tourist arrivals from all the different countries of origin to Catalonia, in [6] to explore effects of swine flu from the effects of earlier global financial crisis on tourism and economy in Brunei. Besides that, Seasonal ARIMA (SARIMA) model was used in [12] to forecast airport passenger traffic for Hong Kong, in [1] to model tourist arrivals and identify the impact of air transport development on the international inbound tourism demand of Saudi Arabia. Both models, ARIMA and SARIMA were employed in [2] to predict tourism expenditure for Belize.

2 Data

To model tourism development in Slovakia are used corresponding monthly unadjusted data of turnover in services of economic activities air transport, accommodation, travel agency, tour operator reservation service and

[1] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, petra.vasanicova@smail.unipo.sk.
[2] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, eva.litavcova@unipo.sk.
[3] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, sylvia.jencova@unipo.sk.

related activities in Slovak Republic by NACE Rev. 2. Data are entered as $index_{2010}$ from January 2000 to January 2017. Dataset was obtained from Eurostat database.

3 Methodology

To purpose the objective of this paper, we model the current and future development of turnover in services of economic activities air transport, accommodation, travel agency, tour operator reservation service and related activities in Slovak Republic by using the exponential smoothing model, and the Box - Jenkins Seasonal Autoregressive Integrated Moving Average (SARIMA) model.

3.1 Holt-Winters multiplicative seasonal model for exponential smoothing

Using exponential smoothing to forecasting time series is the most common. In this paper, we have chosen multiplicative seasonal model on the basis of Holt - Winters' multiplicative method [8, 9]. Forecast model is given as:

$$F_{t+m} = (L_t + b_t m) S_{t-s+m}, \tag{1}$$

where F_{t+m} is the forecast for m periods ahead, s is the length of seasonality, L_t is the level of the series, b_t is the trend, S_t is the seasonal component. Level, trend and seasonal component can be written as:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}), \tag{2}$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}, \tag{3}$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma) S_{t-s}, \tag{4}$$

where Y_t is the observed values. The smoothing parameters α, β, γ, which can be obtain from the range from 0 to 1, were set to fixed values which are determined subjectively by users on the basis of own experience.

3.2 Box-Jenkins SARIMA model

According to [4, 11, 13], a simple equation to define the autoregressive moving average (ARMA)(p, q) model for a stationary time series is given below:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}. \tag{5}$$

The first term in ARIMA model represents an autoregressive (AR) part of the order p having the form of

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t. \tag{6}$$

This (AR) term refers to the current time series values Y_t as a function of past time series values Y_{t-1}, Y_{t-2},..., Y_{t-p}. The ϕ_1, ϕ_2, ϕ_3 are autoregressive coefficients that relates Y_t to Y_{t-1}, Y_{t-2},..., Y_{t-p}.

The moving average MA(q) term of the model is represented as,

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} \tag{7}$$

where, $\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-q}$ are the past random shocks or independent white noise sequence with mean = 0 and variance = σ^2; $\theta_1, \theta_2, ..., \theta_q$ are the moving average coefficients relating Y_t to $\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-q}$.

When the (AR) and (MA) specifications are combined together with integration (differencing) term, they constitute an ARIMA (p,d,q) model, where p, d and q indicate orders of autoregression, differencing and moving average. The model is mathematically given as

$$(1-B)^d Y_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t \tag{8}$$

where, t denotes the time indices, B is the backshift operator, ie., $BY_t = Y_{t-1}$. $\phi(B)$ and $\theta(B)$ are the autoregressive and moving average operators respectively and can be written as

$$\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - ... - \phi_p B^p \tag{9}$$

$$\theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - ... - \theta_q B^q . \tag{10}$$

Seasonality is a pattern which is repeating itself over a fixed time interval. Here, the monthly dataset is presenting a seasonal period of 12 months. In order to obtain a stationarity, seasonal differencing is performed by taking difference between the present and corresponding observation from the previous year. Taking into consideration the seasonality of our time series, a seasonal ARIMA denoted as SARIMA (p, d, q) x $(P, D, Q)_s$, is used, where P,D,Q represent seasonal autoregressive, differencing and moving average orders respectively and *s* is number of seasons. For the present study, $s = 12$. SARIMA(p, d, q)$(P, D, Q)_s$ built for the time series is defined as:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t . \tag{11}$$

where, *B* is the backshift or lag operator, *s* is the seasonal lag (in 'quarter' for present study); ε_t represents error variables; *d* and *D* are non-seasonal and seasonal differences; ϕ and Φ are the non-seasonal and seasonal autoregressive parameters; θ and Θ are the non-seasonal and seasonal moving average parameters respectively.

3.3 Mean absolute percentage error

For determination of validity of our model, we use criterion of the mean absolute percentage error (MAPE). According to [7, 8], MAPE is given by:

$$MAPE = \sum_{i=1}^{n} \left| \frac{(X_i - F_i)/X_i}{n} \right| \cdot 100 . \tag{12}$$

In this equation X_i is the actual data for period *i*, F_i denotes the forecast for period *i*, and *n* is the number of observations. MAPE ≤ 10% means highly accurate forecast, MAPE = 10 – 20% means good forecast, MAPE = 20 – 50% denotes reasonable forecast, and MAPE > 50% denotes inaccurate forecasting.

3.4 Theil's forecast accuracy coefficient

To measure the forecast accuracy we use Theil's inequality coefficient UI, which is given as [3]:

$$UI = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - F_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}X_i^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}F_i^2}} . \tag{13}$$

X_i denotes the actual observation, and F_i denotes corresponding prediction. To measure forecast quality in applied economic forecasting is used modified version of Theil's coefficient UII:

$$UII = \frac{\sqrt{\sum_{i=1}^{n}(F_i - X_i)^2}}{\sqrt{\sum_{i=1}^{n}X_i^2}} . \tag{14}$$

These coefficients is from the interval $\langle 0;1 \rangle$, where UI = 0 means case of equality and UI = 1 means maximum inequality.

4 Results

Using Holt-Winters multiplicative seasonal model for exponential smoothing, the mean absolute percentage error has the smallest value when the parameters $\alpha = 0.795$, $\beta = 0.000$, and $\gamma = 0.000$, and when multiplicative season = 12. In this case, MAPE = 7.415%. According to MAPE interpretation, we can say, that our model

represents highly accurate forecast. Observed, smoothed and forecast values of exponential smoothing model can be seen in Figure 1.

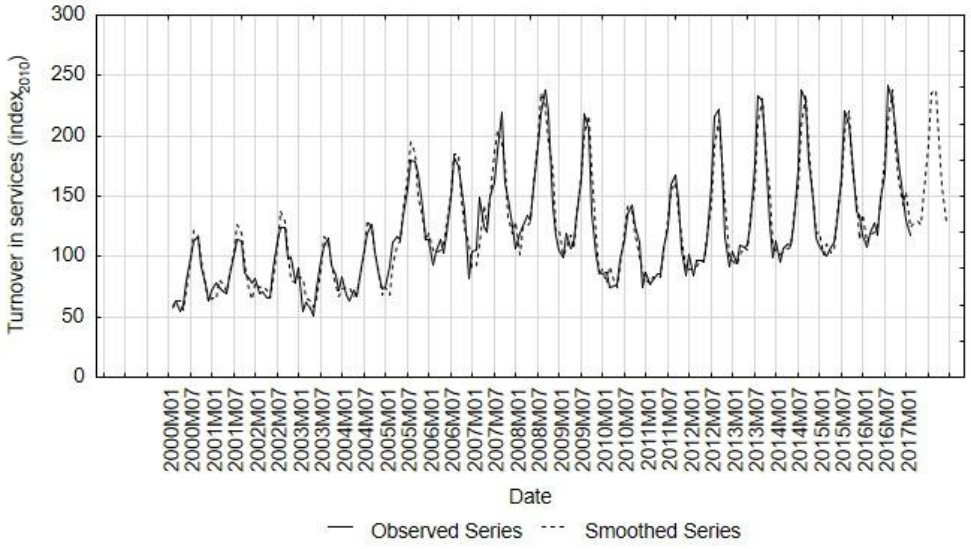


Figure 1 Observed smoothed and forecast series of turnover in services of economic activities connected with tourism using exponential smoothing model

According to SARIMA methodology, the best model is (0,1,1)(0,1,1) with seasonal lag 12. Taking into account assumptions and data fitting, we have chosen the best parsimony model with MAPE value 7.715%, which denotes highly accurate forecast. Obtained parameters for this model are $\theta_1 = 0.317787$, and $\Theta_1 = 0.753040$. Particular model can be written as:

$$X_t - X_{t-1} - X_{t-12} + X_{t-13} = a_t - 0.753040a_{t-12} - 0.317787a_{t-1} + 0.239306a_{t-13}. \quad (15)$$

Observed and forecast series of this model are shown in Figure 2.

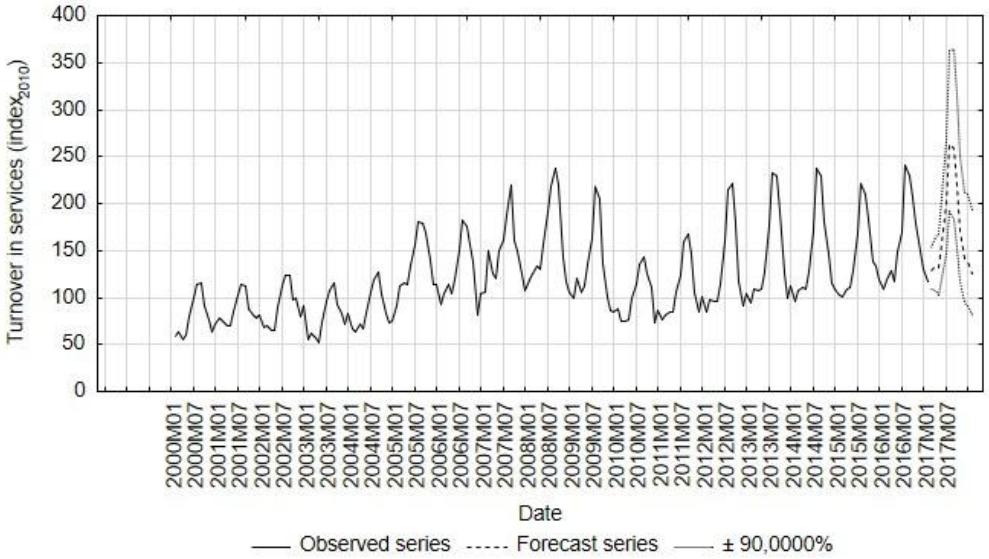


Figure 2 Observed, and forecast series of turnover in services of economic activities connected with tourism using SARIMA model

Forecasted values of turnover in services of economic activities such as air transport, accommodation, travel agency, tour operator reservation service and related activities in Slovakia are presented in Table 1. Even though both models represents high accurate forecast, MAPE value is better for exponential smoothing model. This happens in cases when one analyzes short time series, which are better fitted by using exponential smoothing in compare with ARIMA models. Therefore, exponential smoothing model is more accurate. Compared to observed series, forecast series for corresponds months are increased. Although, values of turnover in services for exponential smoothing model are smaller, we are inclined to this resulting model.

Date	SARIMA	Exponential smoothing
February 2017	129.0533	127.7435
March 2017	133.2830	128.7796
April 2017	131.1195	126.5335
May 2017	157.6648	154.2795
June 2017	198.0086	191.9142
July 2017	263.8492	235.5574
August 2017	259.2478	238.5220
September 2017	219.5696	196.9457
October 2017	171.0927	158.6546
November 2017	141.6907	129.1674
MAPE	7.715%	7.415%

Table 1 Forecast of turnover in services using exponential smoothing and SARIMA model and particular MAPE

Next, we wanted to find out whether predicted values using SARIMA model approximate real observed values. We used shorter time series from January 2000 to December 2015 and we predicted values from January 2016 to April 2017. These predicted values were compared with observed values of turnover in services of economic activities air transport, accommodation, travel agency, tour operator reservation service and related activities in Slovak Republic by NACE Rev. 2. Each observed value of turnover in services is from the confidence interval for forecast value (see Table 2). The value of Theil's forecast accuracy coefficient is UI = 0.0357, and value of modified Theil's coefficient is UII = 0.0719. It means that the SARIMA model seems to be reliable for predicting future values.

Date	Observed value	Forecast value	Lower confidence limit	Upper confidence limit
January 2016	108.5	112.2	90.7	137.1
February 2016	119.1	120.6	93.2	153.7
March 2016	128.0	123.4	91.6	162.7
April 2016	117.6	124.3	89.0	169.0
May 2016	147.5	147.8	102.5	206.5
June 2016	169.0	191.6	129.0	274.5
July 2016	241.5	250.0	163.6	366.4
August 2016	229.9	248.9	158.6	372.8
September 2016	206.3	207.4	128.9	317.0
October 2016	178.7	156.7	95.0	244.1
November 2016	150.0	129.4	76.7	205.3
December 2016	130.2	131.5	76.2	212.3
January 2017	117.6	121.3	67.5	201.7
February 2017	128.0	130.7	70.5	222.5
March 2017	137.6	133.8	70.0	233.0
April 2017	126.1	135.0	68.6	240.1

Table 2 Comparison of observed and forecast values using SARIMA model (0,1,1)(0,1,1)

We have omitted graphs of autocorrelation and partial autocorrelation functions, tests of normality of residuals and many other considered tests and results. However, these results were always taken into account.

5 Conclusion

In this paper was presented modelling of tourism development using the exponential smoothing model, and the Box - Jenkins Seasonal Autoregressive Integrated Moving Average (SARIMA) model. We contributed to existing literature, because we did not model tourism development by using traditional indicators, such as number of arrivals or the nights spent at tourist accommodation establishments, but we used turnover in services of eco-

nomic activities such as air transport, accommodation, travel agency, tour operator reservation service and related activities in Slovakia.

After the comparison of both models was selected the better one, exponential smoothing model, with MAPE value 7.415%, which indicates high accurate forecast. The forecasting values of resulting model are higher than observed values for corresponds months. It means, when turnover in services connected with tourism will increase then development of tourism in Slovakia will increase, too. It follows that tourism in Slovakia has the great potential to contribute to economic growth and growth in employment, especially in the area of economic activities such as air transport, accommodation, travel agency, tour operator reservation service and related activities.

Acknowledgements

Supported by the grant No. 1/0945/17 of the Grant Agency VEGA.

References

[1] Alsumairi, M., and Tsui, K. W. H.: A case study: The impact of low-cost carriers on inbound tourism of Saudi Arabia. *Journal of Air Transport Management* **62** (2017), 129-145.
[2] Banerjee, O., Cicowiez, M., & Cotta, J.: Economics of tourism investment in data scarce countries. *Annals of Tourism Research* **60** (2016), 115-138.
[3] Bliemel, F. : Theil's forecast accuracy coefficient: A clarification. *Journal of Marketing Research* **10** (1973), 444-446.
[4] Box, G. E. P., Jenkins, G. M., and Reinsel, G.C.: *Time Series Analysis: Forecasting and Control*. Prentice-Hall, Englewood Cliff, N.J., 1994.
[5] Claveria, O., and Torra, S.: Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling* **36** (2014), 220-228.
[6] Haque, T. H., and Haque, M. O. (2017). The swine flu and its impacts on tourism in Brunei. *Journal of Hospitality and Tourism Management* (2017), http://dx.doi.org/10.1016/j.jhtm.2016.12.003.
[7] Lee, C. K., Song, H. J., and Mjelde, J. W.: The forecasting of International Expo tourism using quantitative and qualitative techniques. *Tourism Management* **29** (2008), 1084-1098.
[8] Lewis, C. D.: *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann, Borough Green, 1982.
[9] Makridakis, S., Wheelwright, S., and Hyndman, R. J.: *Forecasting: methods and applications*. John Wiley & Sons, USA, 1998.
[10] Stefko, R., Kiralova, A., and Mudrik, M.: Strategic Marketing Communication in Pilgrimage Tourism. *Procedia-Social and Behavioral Sciences* **175** (2015), 423-430.
[11] Taneja, K., Ahmad, S., Ahmad, K., and Attri, S. D.: Time series analysis of aerosol optical depth over New Delhi using Box–Jenkins ARIMA modeling approach. *Atmospheric Pollution Research* **7** (2016), 585-596.
[12] Tsui, W. H. K., Balli, H. O., Gilbey, A., and Gow, H.: Forecasting of Hong Kong airport's passenger throughput. *Tourism Management* **42** (2014), 62-76.
[13] Vasanicova, P., Litavcova, E., and Jencova, S.: On modelling of the development of share of inbound tourism on GDP in Slovakia. In: Conference: 10th International Days of Statistics and Economics (Loster, T., and Pavelka, T., eds.) Prague, CZECH REPUBLIC, 2016, 1938-1947.

Analysis of truncated data with application to the operational risk estimation

Petr Volf[1]

Abstract. Researchers interested in the estimation of operational risk often face problems arising from the structure of available data. The present contribution deals with the problem of left truncation, which means that the values (e.g. the losses) under certain threshold are not reported. Simultaneously, we have to take into account possible occurrence of heavy-tailed distribution of loss values. We recall briefly the methods of incomplete data analysis, then we concentrate to the case of fixed left truncation and parametric models of distribution. The Cramér-von Mises, Anderson-Darling, and the Kolmogorov-Smirnov minimum distance estimators, the maximum likelihood, and the moment estimators are used, their performance is compared, with the aid of randomly generated examples covering also the case of heavy-tailed distribution. Higher robustness of some distance-based estimators is demonstrated. The main objective is to propose a method of statistical analysis and modeling for the distribution of sum of losses over a given period, particularly of its right quantiles.

Keywords: operational risk, severity distribution, truncated data, statistical analysis.

JEL classification: C41, J64
AMS classification: 62N02, 62P25

1 Introduction, the problem of incomplete data

The most traditional field of statistical analysis where the methodology dealing with incomplete data (caused by censoring or truncation) has been developed systematically is the area of statistical survival analysis. While the censoring means that the data values are hidden in known intervals, the truncation arises when some results, though relevant for the analysis, are not reported at all (i.e. we even do not know the number of such lost data). As a rule, there are thresholds (which could be individual and taken as random, or fixed equal for the whole set of observations) such that the values under them (in the case of left truncation) or above them (the right truncation) are not included in available data. It has been shown, for instance already in [7], that when the design of truncation threshold is such that the values from the whole data region are allowed (can be obtained), consistent non-parametric estimation of data distribution is possible. The result has later been extended to regression setting adapting the approach based on counting processes and hazard rate models, the overview is e.g. in [1].

The fixed truncation means that there are no data observed under (or above) a given threshold, therefore only the information on a conditional distribution is available and, in order to fit the complete distribution to such data, its parametric form has to be assumed. The present contribution deals with the case of left truncation. It is inspired by the problem how to estimate the operational risk regulatory capital on the basis of available data-base when the loss data of our interest are truncated from below at a fixed threshold. It is caused by an attempt to avoid recording and storing too many small loss events. However, omitting a part of data makes the problem of modeling operational risk accurately rather difficult [3]. In [6] the authors give an overview of different challenges connected with such an analysis, besides the problem of missing data also the problem of possible heavy-tailed nature of the losses distribution.

The structure of the paper is the following: In the next section the problem will be further specified, the structure of data described, and four methods of losses distribution estimators presented, namely the maximum likelihood estimator (MLE), the moment method (MM), then the Cramér-von Mises (CvM), Anderson-Darling (AD), and Kolmogorov-Smirnov (KS) minimum distance estimators. The methods will be examined on randomly generated data and their performance compared, in particular their reaction to the presence of a part of data coming from a heavy-tailed distribution. It is necessary to emphasize here that the main goal is a reliable estimation (and then the prediction) of the sum of values (losses) over certain period, not only the estimation of parameters of distribution of losses itself. And that the difficulties of analysis are caused principally by two aspects: The truncation (a set of values, though small ones, not recorder at all) and the accidental presence of very high values, outliers from the statistical point of view, which, however, must not be omitted.

[1]Department of Stochastic Informatics, UTIA AV ČR, Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic, volf@utia.cas.cz

2 The problem of heavy tails

In statistics, the robustness of a method (for instance of an estimator) means that its performance is not influenced much by a presence of (a small) portion of outlied values contaminating the regular data. There exists a set of characteristics quantifying the reliability (stability) of a robust method, e.g. the breakdown point or the empirical influence function (cf. [4]). Thus, even in the setting considered here, from the robust statistics point of view, the aim is to estimate well the underlying basic distribution, when it is contaminated by (mixed with) a certain portion of a distribution with heavy tails. To this end, both in [3] and [6] the empirical influence functions for several estimators are derived, showing highly non-robust behavior of the MLE and moment estimators and at least partial robustness of the Cramér-von Mises method (see also [2]). Let us recall here that, in general, a heavy tail of a distribution means that it is not exponentially bounded. In fact, we shall consider here a sub-class of so called "fat-tailed" distributions having its right tail $P(X > x)$ comparable with x^{-a}, for some $a > 0$, as $x \to \infty$.

More specifically, the situation is as follows: We assume that certain parametric distribution type is the baseline model. Further, it is assumed that a realistic model of the data arises from the mixture with another distribution having heavier right tail. In fact, its type can also be specified, still we face a difficult task to estimate parameters of both distributions and the rate of mixture. As the case is further complicated by missing part of data, in general such a problem has no unique solution. Fortunately, in the left truncation case considered here, just certain portion of small values is missing, high values remain available in observed data sets. Then, the main condition of successful model identification is a sufficiently robust method of estimation of the baseline distribution parameters. Hence, the estimators will be compared also from this point of view.

The robustness can be further improved with the aid of convenient robust estimator. In [6] the authors use so called "optimally bias-robust estimator" (OBRE) set of estimators. On the other hand, the structure of left truncated data suggests the use of so called trimmed estimator of the location parameter, i.e. a very simple robust estimation method. That is why we considered such a kind of estimator as a tool for improving the estimation results. However, the improvement was rather negligible, therefore the method is not considered in the follow up.

Proposed estimation procedure has in fact two stages. In the first, the parameters of the baseline distribution are estimated. To do it reliably, sufficiently robust estimator should be employed. Then, on the basis of well estimated parameters of the baseline distribution, the second component of the mixture and the mixture rate can be estimated, which is crucial for the main goal of the analysis, namely for prediction of aggregated losses. This stage, on the contrary, has to use an estimator sensitive to all values, in order to distinguish both mixture components.

In the sequel we shall consider, similarly as [3] and [6], the log-normal baseline distribution of losses, as it is a model convenient both from practical and theoretical point of view. Further, its right part will be contaminated by the Pareto distribution as a model of possible occurrence of large values, as it is commonly considered to be a reasonable choice [5]. Again, let us recall here briefly that the Pareto (or also "power law") distribution has distribution function $F_p(x) = 1 - (A/x)^\lambda$ for $x > A > 0$, $F_p(x) = 0$ for $x \leq A$, $\lambda > 0$ is its shape parameter.

3 The model and estimators

It is assumed that a positive random variable X is observed just when its value is above a given threshold T. Hence, the data consist of a random sample $X_i, i = 1, .., N_1$, all $X_i > T$. The part under T is not observed, nor its frequency N_2 is known. Denote the density function of X $f(x)$, distribution function $F(x)$. It is further assumed that this distribution is a mixture, namely $f(x) = (1 - \alpha) \cdot f_0(x) + \alpha \cdot f_1(x)$, where the basic part $f_0(x)$ is given by a log-normal distribution with unknown parameters μ_0, σ_0, and is contaminated by a Pareto distribution, with density function $f_1(x)$ and with appropriate parameters. As it has been said above, both its parameters and the rate of contamination are also the object of estimation. We assume that the contamination rate α is not large, we have examined its influence for $\alpha \in [0,\ 0.2]$. Thus, the first goal is to estimate parameters of $f_0(x)$. As it has been said above, the aim of this first stage is to use a sufficiently robust procedure. Just for comparison, we shall deal with cases both with and without contamination, examining the behavior of several estimators. Namely the MLE, moment estimator and three distance-based methods.

Remark: The assumption of log-normal distribution allows to work with normal distribution model for logarithmized data. Hence the methods described above can be used for transformed data, it can simplify numerical procedures. As regards the contamination, let us recall that logarithmized Pareto distribution yields the exponential one. This connection will be used later for the random generation of data.

3.1 Estimation methods

In the case of full data, we can construct the full empirical distribution function, as a reliable non-parametric distribution estimate. Under the assumption of parametrized distribution, let us denote its density $f(x; \theta)$, distribution

function $F(x;\theta)$, set of parameters hidden in θ should be estimated. From the fixed truncation it follows that the part of distribution above threshold T is given by the density and distribution functions, resp., both for $x > T$:

$$f_T(x;\theta) = \frac{f(x;\theta)}{1 - F(T;\theta)}, \quad F_T(x;\theta) = \frac{F(x;\theta) - F(T;\theta)}{1 - F(T;\theta)}.$$

1. Maximum likelihood estimator. The likelihood based on observed data has the form

$$L(\theta, \boldsymbol{x}) = \prod_{i=1}^{N_1} f_T(X_i;\theta)$$

and we search for θ maximizing its logarithm.

2. Moment estimator. Let us compute conditional first 2 moments of $(X|X > T)$ and compare them with empirical moments obtained from observed data. Namely, we shall compute

$$E_T^{(k)}(\theta) = \int_T^\infty x^k \, f_T(x;\theta)\, dx, \quad \bar{X^k} = \frac{1}{N_1} \sum_{i=1}^{N_1} X_i^k.$$

The best θ should minimize a distance of them, in the simplest case $\sum_{k=1}^{2}(E_T^{(k)}(\theta) - \bar{X^k})^2$.

3. Cramér-von Mises estimator. It minimizes the distance between the empirical and assumed distribution function on (T, ∞), namely we search for θ minimizing

$$\sum_{i=1}^{N_1}(F_{emp,T}(X_i) \; - \; F_T(X_i;\theta))^2,$$

where $F_{emp,T}(x)$ is the empirical distribution function computed from data observed above T. Namely, the simplest form is $F_{emp,T}(X_{(i)}) = i/N_1$, $i = 1, ..., N_1$, where $X_{(1)} \leq X_{(2)} \leq, ..., \leq X_{(N_1)}$ denote ordered observations. We use the following variant: $F_{emp,T}(X_{(i)}) = (2i-1)/2N_1$.

4. Anderson-Darling estimator is a weighted variant of the CvM estimator giving the data-points weights corresponding to the variance of empirical distribution function. Hence, it minimizes

$$\sum_{i=1}^{N_1}(F_{emp,T}(X_i) \; - \; F_T(X_i;\theta))^2 \cdot \frac{1}{w_i},$$

where $w_i = F_T(X_i;\theta) \cdot (1 - F_T(X_i;\theta))$. The weighting results in a higher sensitivity to small and large data, hence also in smaller robustness compared to the CvM method. However, still its influence function is bounded. This difference actually will lead us to the estimator choice, on the basis of following Monte Carlo study. In the first stage, where rare outlying data should have small influence, the CvM estimator will be preferred. Further, however, when the model should describe well also the source of contamination, the AD estimator will be utilized.

5. Kolmogorov-Smirnov estimator is based on minimizing the maximal distance between empirical and model distribution functions, i.e. it minimizes

$$\max_{X_i} |F_{emp,T}(X_i) \; - \; F_T(X_i;\theta)|.$$

It is evident that in all cases the estimation has to be solved with the aid of a convenient numerical optimization procedure, the moment method evaluation includes also numerical integration.

4 Monte Carlo study

The study is based on K-times repeated generation of data sets of extent N. Each such set is taken as representing the loss data over certain period. The data have been generated from normal distribution with parameters μ_0, σ_0 and mixed with values from exponential distribution with parameter λ shifted by a constant a, i.e. having distribution function $F_e(x) = 1 - \exp(-\lambda \cdot (x - a))$ for $x \geq a$. The mixture (contamination) rate α was selected from $[0, 0.3]$. Such data represented logarithms of losses, they then were truncated from the left side by a threshold T_0. Hence, the losses were given by values coming from the mixture of log-normal distribution (with μ_0 and σ_0) with the Pareto distribution having distribution function $F_p(x) = 1 - (A/x)^\lambda$ for $x \geq A = \exp(a)$. The values of losses were truncated by threshold $T = \exp(T_0)$.

The set of truncated data then contained just $N_1 \leq N$ values greater than the threshold, it was assumed that the number of omitted data as well as their values were not known. In fact, as the data were prepared artificially, we knew them and could use them as a benchmark for comparison of performance of estimation methods and examination of information loss caused by the truncation. As in each Monte Carlo study, the repetition of analysis enabled us to construct empirical distribution of estimates, to study their bias and variability, and, later on, to analyze and compare distributions of sums reconstructed on the basis of different estimation methods.

The example provided here uses the following values: $\mu_0 = 2$, $\sigma_0 = 0.5$, $\lambda = 1$, $a = 2$, hence $A \doteq 7.39$. Further $T_0 = 1.3$, $\alpha = 0$ or 0.1, $N = 1000$, $K = 1000$ were selected. From such a choice it follows that the basic log-normal distribution had expectation $\doteq 8.4$ and standard deviation $\doteq 4.5$, while the Pareto distribution with parameter $\lambda = 1$ had infinite all moments. Threshold $T = \exp(1.3) \doteq 3.67$, the proportion of data truncated off was about 8%. Just for comparison, the 95% quantiles were 16.8 and 147.8 for these log-normal and Pareto distributions, respectively, 99% quantiles were 23.6 and 738.9.

4.1 Results of parameters estimation

The first case examined was the case without contamination, the data were generated just to correspond the log-normal distribution with given parameters μ_0, σ_0. Data were then truncated and parameters estimated from truncated samples by three methods. As the data generation was repeated K times, K estimates were obtained for each parameter and each method. Figure 1 displays these sets of estimates in a form of boxplots. The first correspond to the MLE from complete data, the other three then to the CvM estimator, the MLE and to moment estimator. It is seen that their performance is comparable, bias negligible and variability increased (compared to estimates from full data) due a loss of information caused by the truncation. Other estimators (KS and AD) performed very similarly.

In the second case presented here the log-normal (μ_0, σ_0) data were mixed with values generated from the Pareto distribution, their proportion was $\alpha = 0.1$. As it was said, during this stage of analysis the data were still treated as coming from log-normal distribution with unknown parameters μ, σ. Figure 2 again shows the results of estimation, in K repetitions, first the MLE from full data, then the results of 3 selected estimation methods used to truncated data. Now the pattern is different. First, as the contamination has caused a number of large, outlying values in data, the consequence is that the estimates are shifted, namely estimated standard deviation is increased and estimate of μ biased even in the case of the MLE from full data. Further, reactions of examined estimation methods to contaminated and truncated data differ. As expected, both the MLE and moment estimators react by even more increased both bias and variability of values, relative to estimates obtained from full data. On the other hand, in order to cope with heavier right tail of the data, the CvM method yielded a slightly increased estimates of both μ andσ. Simultaneously, variability of estimates did not increase significantly, which indicates a consistency of method. Such an phenomenon can be related to findings in [6] concluding that the CvM method is much more robust (having bounded empirical distribution function) than the other two. Further, as regards the other distance-based estimators, the result is collected in Table 1. It is seen that the KS method yielded results quite comparable with those of the CvM, while the AD estimators showed a stronger reaction to right tail data, it was biased and had larger variability similarly like the MLE and the moment method.

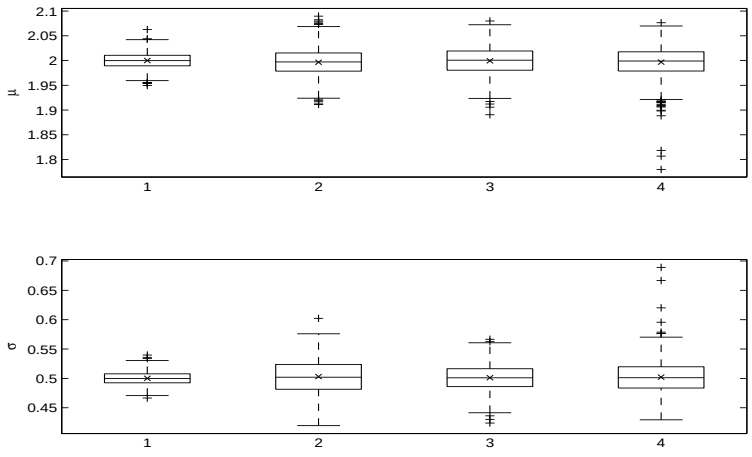


Figure 1 Estimated μ (above) and σ (below) in the case of no contamination: 1–MLE estimates from complete data, 2–CvM estimator, 3–MLE, 4–moment method.

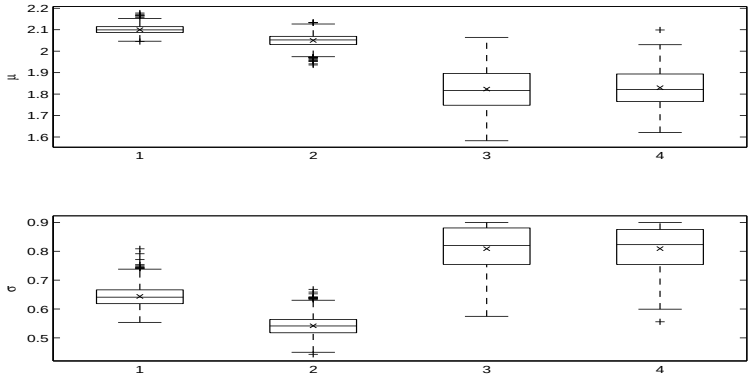


Figure 2 Estimated μ (above) and σ (below) when contamination rate was $\alpha = 0.1$: 1–MLE estimates from complete data, 2–CvM estimator, 3–MLE, 4–moment method.

	estimated: μ				estimated: σ			
Method	mean	median	Q(0.05)	Q(0.95)	mean	median	Q(0.05)	Q(0.95)
CvM	2.0493	2.0512	1.9952	2.0954	0.5431	0.5416	0.4900	0.6007
KS	2.0441	2.0460	1.9841	2.0966	0.5533	0.5529	0.4930	0.6195
AD	2.2426	2.1097	1.7212	2.9800	0.8682	0.8918	0.7494	0.8995
MLE	1.8230	1.8132	1.6812	1.9806	0.8073	0.8215	0.6708	0.8963
Moment	1.8298	1.8219	1.6984	1.9778	0.8092	0.8230	0.6749	0.8952

Table 1 Empirical characteristics of estimates obtained from different methods.

4.2 Analysis of contamination

In the second estimation stage the aim is to identify the heavy-tailed component of the mixture and estimate its parameters, when the Pareto model is assumed.Hence, the method should be sensitive to all observed values, giving an appropriate weights also to right tails of data. After a set of experiments we decided to prefer the AD estimator meeting best such requirements. The numerical example presented here, again based on K sets of N data (partly left-truncated), and using μ andσ estimated in the first stage, yielded the estimates which empirical characteristics (from K repetitions) are summarized in Table 2.

Parameter	mean	median	Q(0.05)	Q(0.95)
a	1.695	1.521	1.321	2.982
λ	0.839	0.903	0.121	1.602
α	0.094	0.091	0.014	0.217

Table 2 Empirical characteristics of estimates.

It is seen that empirical distribution of estimates is not symmetric, still rather wide, but at least the mean or median values providing acceptable results. Simultaneously, certain trade-off among parameters can be traced. For instance, smaller λ leads to longer right tail, while smaller a shifts the whole distribution left.

4.3 Estimated distribution of sums

As it has been said, this task is the main and final objective of the study. In particular, we are interested in how well the methods are able to model (and then to predict) upper right end quantiles of distribution of sums. This distribution is very sensitive even to just small changes of parameters, hence also to their unperfect estimates. And we have seen how rather complicated the estimation procedure is. Simultaneously, the results depends also on the number of losses during given period. This point is not considered here, we just try to estimate the distribution of convolution of a fixed number, D, of i.i.d. random variables representing the losses. The recommended approach to the operational risk modeling concerns the calculation of a risk measure VaR_γ at a confidence level $\gamma = 99,9\%$ for a loss random variable L corresponding to the aggregate losses over a given period, usually one year [5]. As this distribution has no closed form, standard way of examining it is again a Monte Carlo approach. Therefore we generated K times, with K 10^5, sums $L = \sum L_k$ of $D = 100$ variables L_k having the mixed distribution derived and estimated in preceding parts. Table 2 shows a comparison of chosen right empirical quantiles of L obtained

by random generation.

Quantile	0.995	0.996	0.997	0.998	0.999	0.9995
a) Estimated	18 730	23 671	31 638	46 747	97 711	202 301
b) "True"	15 555	19 183	25 214	37 952	84 846	154 176

Table 3 Empirical quantiles of L based on a) model using the medians of estimated parameters $\mu = 2.0512, \sigma = 0.5416, \lambda = 0.903, a = 1.521, \alpha = 0.091$; b) the "true" model with parameters $\mu_0 = 2, \sigma_0 = 0,5, \lambda_0 = 1, a_0 = 2, \alpha_0 = 0.1$.

The quantiles based on estimated parameters exceed slightly the quantiles of true distribution of sums. It indicates that the method could be applicable without large danger of underestimation of real aggregate losses. Naturally, each analysis of this kind has to start from careful exploration of available real data.

5 Concluding remarks

The first aim of the study was to examine and compare performance of several estimators of distribution parameters in the case of fixed left truncated data. The data were generated randomly, the sense of examples was to simulate a set of losses of a financial institution encountered during certain period. Their distribution was modeled via the log-normal distribution contaminated by the Pareto one. The main objective was then the estimation of distribution of sums of losses over a given period..It means to summarize the values coming from (possibly contaminated) log-normal distribution and, moreover, not observed fully. Theoretically, the distribution could be approximated on the basis of the central limit theorem. However, there are many issues leading to doubts on its correctness and practical usefulness. The asymptotic behavior of the C.L.Th. on distribution tails is rather slow in general, not speaking about the fact that Pareto distribution of our choice does not fulfil theoretical requirement for the C.L.Th. validity.

That is why this part of analysis was also based on Monte Carlo approach and estimated parameters. We hope that such an approach is suitable also for practical use. As a rule, a sufficiently large database is available, usually omitting values under given threshold. Hence, the parameters of assumed type of baseline distribution can be estimated, e.g. using sufficiently robust Cramér-von Mises estimator. Then the model for heavy-tailed part of losses distribution can be identified, in this stage a less robust method is appropriate, we can recommend the Anderson-Darling method. Finally, random generation from obtained model helps to recover expected behavior of aggregated losses.

References

[1] Andersen, P. K., and Keiding, N.: *Survival and Event History Analysis*. John Wiley & Sons, New York, 2007.

[2] Duchesne, T., Rioux, J., and Luong, A.: Minimum Cramér-von Mises estimators and their influence function. *Actuarial Research Clearing House* **1** (1997), 349–361.

[3] Ergashev, B., Pavlikov, K., Uryasev, S., and Sekeris, E.: Estimation of truncated data samples in operational risk modeling. *The Journal of Risk and Insurance* **83** (2016), 613-640.

[4] Huber, P. J., and Ronchetti, E.: *Robust Statistics (2-nd Edition)*. John Wiley & Sons, New York, 2009.

[5] Nešlehová, J., Embrechts, P., and Chavez-Demoulin, V.: Infinite mean models and the LDA for operational risk. *Journal of Operational Risk* **1** (2006), 3–25.

[6] Opdyke, J. D., and Cavallo, A.: Estimating operational risk capital: the challenges of truncation, the hazards of maximum likelihood estimation, and the promise of robust statistics. *Journal of Operational Risk* **7** (2012), 3–90.

[7] Turnbull, B. C.: The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)* **38** (1976), 290–295.

Implementation of permutation tests in research of problematic use of the Internet by young people

Katarzyna Warzecha[1], Tomasz Żądło[2]

Abstract. The Internet is for many young people the source of entertainment and acquiring new knowledge which contributes to their development and enables interactive education. However, on the other hand the problematic use of the Internet becomes more and more of psychosocial problem. The incorrect use of information and communication technologies may have negative influence on mental, physical and emotional development of young people and the loss of control over the Internet use may result in neglecting the schoolwork, the lack of physical activity and worsening of the relations with other young people and with family. The main purpose of this paper is the study of middle school young people in Silesian voivodship in Poland in terms of risky Internet behaviours and typical Internet acitivies using survey sampling methods and permutation tests.

Keywords: permutation tests**,** Internet addiction, young people.

JEL Classification: C12, C13, C83
AMS Classification: 62F03, 62F10, 62D05

1 Introduction

In contemporary world the information and communication technologies (ICT) recognized as computers, tablet computers or mobile phones frequently with Internet access are used by young people and it is considered to be the massive and common behavior. More and more common access to the Internet has got many positive aspects (the quick accessibility of information and knowledge, instant processing and transmission of data and unlimited and express communication which helps to keep the relations with people and obtain new contacts). On the other hand problematic use of the Internet becomes more and more of psychosocial problem.

The use of ICT is considered to be a phenomenon to which one can become dependent on and which can have negative impact on mental, physical and emotional development of an individual [9]. The main purpose of the survey was the estimation of fraction of middle school young people in Silesian voivodship who problematically use the Internet based on IAT Kimberly Young test. Two research questions considered in the paper are defined as follows:

- Is there a dependence between different types of Internet activities and type of Internet usage (problematic/unproblematic)?
- Is there a dependence between sex of respondents and type of Internet usage (problematic/unproblematic)?

The calculations will be made with the use of R Cran and Microsoft Excel programs. The sample was drawn at random from the population of middle school young people in Silesian voivodship in Poland using multistage stratified sampling.

1.1 Permutation tests

From the population of middle schools' students Ω of size $N = 116\,460$ we draw at random a sample s of size $n = 1598$ using multistage stratified sampling design denoted by $P(s)$ with first and second order inclusion probabilities denoted by π_i and π_{ij}, respectively. The set of respondents denoted by s_r of size denoted by n_r is a subset of s due to the nonresponse – in our study $n_r = 1333$. We consider the following calibration estimator of the mean in the subpopulation Ω_c of size N_c (see [16] p. 58):

$$\hat{\bar{Y}}_c = N_c^{-1} \sum_{i \in s_r} a_{ic} w_i y_i \,, \tag{1}$$

[1] University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, 50, 1 Maja Street, 40-287 Katowice, Poland, e-mail:katarzyna.warzecha@ue.katowice.pl
[2] University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, 50, 1 Maja Street, 40-287 Katowice, Poland, e-mail:tomasz.zadlo@ue.katowice.pl

where $a_{ic}=1$ if $i\in\Omega_c$ and 0 otherwise, $w_i=d_i v_i$, $d_i=\pi_i^{-1}$, $v_i=1+\lambda^T\mathbf{x}_i$, $\lambda^T=\left(\mathbf{X}-\sum_{i\in s_r}d_i\mathbf{x}_i\right)^T\left(\sum_{i\in s_r}d_i\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}$, $\mathbf{x}_i$ is a vector of auxiliary variables for *i*th element in the sample, **X** is the vector of population totals of auxiliary variables. The estimator of variance of (1) is given by (see [16] p. 136):

$$\hat{D}^2(\hat{\bar{Y}}_c)=\hat{V}_{SAM}+\hat{V}_{NR}, \tag{2}$$

where the component resulting from the sampling design distribution is as follows:

$$\hat{V}_{SAM}=N_c^{-2}\sum_{i\in s_r}\sum_{j\in s_r}a_{ic}(d_i d_j-d_{ij})(v_i\hat{e}_i)(v_j\hat{e}_j)-N_c^{-2}\sum_{i\in s_r}a_{ic}d_i(d_i-1)v_i(v_i-1)(\hat{e}_i)^2, \tag{3}$$

and the component resulting from the nonresponse distribution equals:

$$\hat{V}_{NR}=N_c^{-2}\sum_{i\in s_r}a_{ic}v_i(v_i-1)(d_i\hat{e}_i)^2, \tag{4}$$

where $\hat{e}_i=y_i-\mathbf{x}_i^T\mathbf{B}$, $\mathbf{B}=\left(\sum_{i\in s_r}d_i v_i\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\sum_{i\in s_r}d_i v_i\mathbf{x}_i y_i\right)$. As auxiliary variables we use dummy variables defining the type of the school, the size of the city or village, the age of students and their sex.

To test equality of means we are not able to use classic tests known from mathematical statistics because the sample is not simple random sample and because of the nonresponse. Even in the case of the lack of the nonresponse the problem is complex and usually (see [2] pp. 137-139) the normality of the distribution of a test statistic is assumed. To avoid these problems we use the class of permutation tests discussed in e.g. [15] and [10]. We consider some test statistic denoted by T, where high values of T allow to reject the null hypothesis. In the case of testing equality of means we consider two vectors: vector $\mathbf{y}$ of the variable of interest and vector $\mathbf{z}$ of identifiers of subpopulations in the sample. If the number of all permutations of the sample data is too large, the following algorithm is used ([15] p. 45):

1. Based on real sample data set we compute the value of the test statistic denoted by $T_0=T(\mathbf{y},\mathbf{z})$.
2. We permute elements of vector **z** and obtain vector $\mathbf{z}^*$.
3. We compute the value of the test statistic $T^*=T(\mathbf{y},\mathbf{z}^*)$.
4. We repeat steps 2 and 3 *B* times to obtain *B* values of $T^{*b}=T(\mathbf{y},\mathbf{z}^{*b})$, where $b=1,2,...,B$.
5. We estimate p-value as $B^{-1}\sum_{1\le b\le B}I(T^{*b}\ge T_0)$.

If B diverges to infinity, the p-value strongly converge to the true value ([15] p.45). We assume $B=10000$. For the test where $H_0: m_1=m_2$ and $H_1: m_1>m_2$ we use the following test statistic:

$$T=\hat{\bar{Y}}_i-\hat{\bar{Y}}_j, \tag{5}$$

where $\hat{\bar{Y}}_i$ and $\hat{\bar{Y}}_j$ are given by (1).

In the case of complex samples and the lack of nonresponse, to test independence of two categorical variables we can use the classic Chi-square test statistic corrected by some constant which depends on the sampling design (see [2] p. 149). In our study, due to the nonresponse we use permutation version of the Chi-square test, where the test statistic is given by the classic Chi-square statistic:

$$\chi^2(\mathbf{x},\mathbf{y})=\sum_{i=1}^{k}\sum_{j=1}^{l}\hat{n}_{ij}^{-1}(n_{ij}-\hat{n}_{ij})^2 \tag{6}$$

where **x** and **y** are vectors of sampled values of *k*-way and *l*-way categorical variables, k and l are number of rows and columns of the contingency table, respectively, n_{ij} and $\hat{n}_{ij}$ are real and expected counts in *i*th row and *j*th column of the contingency table, respectively. It means that in the step 1 of the procedure presented above $T_0=T(.)$ is replaced by $T_0=\chi^2(\mathbf{x},\mathbf{y})$, where χ^2 is given by (6). Moreover, in steps 3 and 4 statistics $T^*=T(.)$ and $T^{*b}=T(.)$ are replaced by $T^*=\chi^2(\mathbf{x},\mathbf{y}^*)$ and $T^{*b}=\chi^2(\mathbf{x},\mathbf{y}^{*b})$, respectively, where $\mathbf{y}^*$ is a permuted vector $\mathbf{y}$ and $\mathbf{y}^{*b}$ is a permuted vector $\mathbf{y}$ in the *b*th iteration.

1.2 Problematic use of the Internet

Problematic use of the Internet by adolescents and young adults becomes more and more of psychosocial problem. Making use of information and communication technologies (ICT) is considered to be a phenomenon which have negative impact on mental, physical and emotional development of an individual. The core of this phenomenon became significant in recent years mainly for psychiatrists, psychologists, educators, sociologists

and addiction therapists who adopt various theoretical and empirical foundations by explaining and analyzing the addictions [8].

In the source literature the dysfunctional use of the Internet is associated with many terms and despite the fact that this phenomenon is treated with great measure there are many notions attached to one definition [1; 7] including among others: Internet addiction [4]; problematic Internet use [3]; pathological Internet use [13]; Internet abuse [12], dysfunctional Internet use [1]. The main issue of this research is risky / problematic Internet use and it is due to the fact that middle school students who have been the subject of the research are the people who have just begun their independent, intensive contact with the Internet and with the modern means of communication such as tablet computer and smartphone with Internet access.

In the research, the test of problematic Internet use was implemented (Internet Addiction Test by Young, [18]; Polish adaptation by Hawi, Błachnio, Przepiórka [6]). This test is based on the criteria of habits control and drives disorders according to DSM-IV. The tool used in the research has got optimal psychometric characteristics Crobach's α=0,93, which were tested on the sample of N=6119 persons. The problematic Internet use is understood as the disorder of control over impulses and the notion of Internet relates to all of the on-line activities. The test measures how the Internet influences various aspects of everyday life (e.g.: everyday activity, pattern of sleep, productivity, social life, emotions).

It was agreed to make use of the above test as a research tool due to its popularity and reliability (the described scale has got confirmed high psychometric parameters). The obtained results will enable evaluation of the tested phenomenon range and they will be compared with other national and foreign test results. As it is presented in the research conducted by many science centers the problematic Internet use varies according to the country in which the research was conducted and the research tool which was used [5]. Polish research made by Young's test indicate that 2,8% problematically use the Internet [14] and that the risk of and problematic Internet use itself relate more often to boys than girls [14;11].

2 Identification of problematic Internet use phenomenon in the group of middle school students

This article is a presentation of a small part of empirical research results which were conducted between 2015 and 2016 on the representative sample of middle school adolescents in Silesia voivodship. The conducted research focused on the threats connected with the use and overuse of the new media (the Internet, mobile phones, computer games and gambling) by adolescents. The questionnaire for students was constructed by the team realizing the research in cooperation with the specialists of addiction therapy and it contained the screen tests to research chosen behavioral addictions (screen tests of pathological Internet use (Young's test- IAT: Internet Addiction Test; SOGS-RA (The South Oaks Gambling screen Revised Adolescent: an instrument for the identification of pathological gamblers); AICA-S test (Assessment of Internet and Computer game Addiction) as the diagnostic tool to research the addiction to computer games). We use multistage stratified sampling design where strata were definied based on the type of the school and the size of the city. Students within sampled schools were additionally stratified due to their age. The sampling frame was the register of schools received from Silesian Board of Education. The population size was 116460 and the sample size – 1333. The implementation of the research was co-financed from the resources of Gambling Problem Solving Fund in the framework of Program I of Ministry of Health.

Internet activities of middle school adolescents

The results of the research conducted on Silesian adolescents indicate that middle school students are quite well equipped in modern means of communication. It is estimated that in the population of middle school students in Silesian voivodship the desktop computer (one or more than one) was possessed at home by 66% of middle students, slightly over the half of middle school students had laptop computers (52%), approximately 35% of middle school students owned a mobile phone and approx. 58% owned a smartphone. More and more popular among adolescents, tablet computers belonged to 65% of middle school students. It is estimated that among Silesian students population, Internet connection at homes was possessed by 90% of middle school students and 75% of them owned the Internet access in their phones. Silesian middle schools are also well equipped in computer labs and the students have the access to computers with Internet connection. Over a span of researched 10 years the accessibility to Internet in Silesian schools among adolescents significantly improved (very poor conditions are invariably in Sosnowiec, Ruda Śląska, Jaworzo and Zabrze, the best conditions are in częstochowski and gliwicki districts and in Piekary Śląskie. Częstochowski district is best equipped in computer hardware and there are about 9 middle school students for 1 computer with Internet connection and 88.24% of middle schools are equipped with the computers with Internet access for students to use. The worst equipment in computer hardware in schools is noticed in Jastrzębie Zdrój district. There are about 17 students for 1 computer

with Internet connection in middle schools and 75% of middle schools are equipped with the computers with Internet access for students to use [17].

The most frequent Internet activities among tested middle school adolescents can be found below (the research over activity and the estimated percentage is presented in Table 1).

Types of Internet activities	**Percentage of middle school students**	**The permutation version of the Chi-square test – below given p value**	
		3 groups isolated according to IAT	**sex**
Checking e-mails	82.11	0.351	0.101
Participation in online chats	84.13	0.000	0.002
Web browsing	94.88	0.002	0.457
Collecting materials for learning	95.51	0.001	0.000
Buying products in the Internet	74.78	0.000	0.007
Participation in Internet auctions	47.81	0.000	0.000
Playing online games	76.30	0.000	0.000
Downloading free software	58.88	0.000	0.000
Downloading free music, films	82.93	0.000	0.000
Obtaining information from websites	89.22	0.137	0.731
Listening to music / radio online	88.41	0.002	0.001
Watching TV / films online	92.26	0.570	0.000
Booking tickets (cinema / match)	56.65	0.000	0.003
Checking information in e.g. online register	74.28	0.000	0.000
Participation in social networking	92.92	0.000	0.000
Reading newspapers / articles online	69.24	0.238	0.005
Participation in drawings / online bets	16.82	0.000	0.000
Browsing websites with pornographic content	23.77	0.000	0.000
Flirting on online chats	36.60	0.000	0.000
Using Snapchat	50.21	0.000	0.000
Making selfies and uploading them in the Internet	64.76	0.000	0.000
Shooting short films by phones and in uploading them the Internet	42.28	0.000	0.034

Numbers in gray color are the most frequent activities made with the use of computer.

Table 1. Estimated fraction of Silesian middle school students according to types of online activities and the permutation version of the Chi-square test results

It is estimated that Silesian middle school students most frequently use the Internet to:

- participate in social networking (60% participate in social networking every day or almost every day);
- browse websites (42% browse websites every day or almost every day);
- listen to music / radio online (44% listen to music / radio online every day or almost every day);
- search educational information in order to prepare for the lessons (18% search such information every day or almost every day);
- watch TV / films online (every fifth student watches TV or film online every day or almost every day);
- download and listen to music (36% listen to music online every day or almost every day).

Some of the online activities of students are alarming: from the estimated fraction of Silesian middle school students as regards their online activities it is noticeable that about 17% of them gamble online by drawings, bets and casinos and every fourth middle school student watches websites with pornographic content. The time spent online by Silesian students differentiates according to the day of the week (weekday or weekend). From the data, it can be observed that among Silesian middle school students every fifth student spends half an hour online every day on weekdays and at weekends as well, while 22% of students spend more than 5 hours online everyday on weekdays as well as at weekends.

Testing hypotheses – Independence test – Permutation version of chi-squared test

One of the purposes of the research was examination whether there are significant dependencies in intensity and purposes of using the Internet between the adolescents using the Internet correctly and between those who use it pathologically / problematically (Table 1). Main conclusions based on permutation Chi-square test presented in Table 1 are:

- the is no dependence between sex and the following Internet activities: checking e-mails; web browsing; obtaining information from websites
- the is no dependence between the type of Internet usage (3 groups of students isolated according to IAT: using the Internet in a correct way, at risk of using the Internet problematically, problematically using the Internet) and the following Internet activities: checking emails, obtaining information from websites, watching TV, films online and reading newspapers, articles online

Problematic use of the Internet – results of Young's test – Internet Addiction Test IAT

Problematic use of the Internet was measured with Young's Internet Addiction Test IAT which is a test to measure the overuse of this kind of entertainment. It enabled to estimate the risk scale of problematic Internet use among Silesian middle school students. It is estimated that (Table 2):

- 1.49% of Silesian middle school students problematically use the Internet, where more boys than girls use the Internet problematically;
- 8.88% of Silesian middle school students are at risk of using the Internet problematically, where more boys than girls are at this risk;
- unquestionable majority – 89.96% of Silesian middle school students – are people using the Internet in a correct way.

The results of research as regards the estimation of subpopulation characteristics with the mean errors of estimation and the verification of applied hypotheses (a permutation test was used - a test of equality of two means in subpopulations of boys and girls) are presented in Table 2. The estimated mean value of this test in middle school students community equals to 27.5434 with estimated standard terror of estimation equal to 0.7091. Hence, the estimated relative standard error of estimation is less than 2.6%. What is important here, on the basis of results presented in Table 2 it can be stated that mean Young's IAT test result is significantly higher in the group of boys than in the group of girls.

estimated parameter / population/ subpopulation	overall mean test result	Fraction of students using the Internet in a correct way	Fraction of students at risk of using the Internet problematically	Fraction of students problematically using the Internet
middle school students	**27.5434 (0.7091)**	**0.8963 (0.0110)**	**0.0888 (0.0083)**	**0.0149 (0.0051)**
girls	26.5413[(a)] (0.6425)	0.9075 (0.0122)	0.0788 (0.0108)	0.0137 (0.0047)
boys	28.4849[(a)] (1.1655)	0.8859 (0.0141)	0.0981 (0.0125)	0.0160 (0.0093)

(a) With 0.95 probability it can be stated that the mean in subpopulation of boys is higher than mean in subpopulation of girls (p=0.0272).

Table 2. Estimation of middle school students subpopulation characteristic on the basis of Young's Internet Addiction Test IAT results (in brackets: values of estimated standard errors) and the result of the permutation test (the test of equality of two means in subpopulations of boys and girls)

3 Conclusions

The conducted research shows that as a result of a rapid technological development adolescents are more and more receptive to have modern means of communication (in particular the devices with Internet access), and the facility of the Internet access and the wide range of services offered inside may have the influence on problematic use of the Internet or playing computer games.

As the research shows, the dominant group of Internet users is the group using this medium in a correct way. For them, it is a new technical improvement helping in obtaining knowledge and widening the mental horizons. The next group among researched Silesian adolescents were the people who are losing control over using the Internet as regards time, the way of use in general and for computer games, as well as neglecting their everyday duties. The Internet becomes their natural environment of existence as a result of spending many hours in front of the computer with Internet access- this may result in health problems and damages in various areas of activities as well as disorganization of functioning in real life. Thus, the incorrect way of using the Internet is connected with negative consequences that may have the influence on their health. Problematic Internet use and being at risk of problematic Internet use concerns more the group of boys (1.6% and 9.81%) than the group of

girls (1.37% and 7.88% respectively) among students of middle schools. Moreover, the estimated percentage of Silesian middle school students who problematically use the Internet equals to 1.49%.

The results of the research contributed to filling the gap in the status of knowledge as regards the problematic Internet use and online activities among Silesian adolescents. The permutation tests that were used are a useful tool especially when the research is not conducted on simple sampling which prevents the implementation of classic procedures of hypotheses testing verification such as equality of means and independencetests known from the mathematical statistics.

Acknowledgements

The research was conducted in 2015-2016 in the framework of Ministry of Health's grant (contract no. 67/HE/2015 and 1/HEK/2016) Program I: Supporting scientific research concerning the phenomenon of gambling and other behavioral addictions and related problems. The proper name of program: *The profile of gambling phenomenon and pathological Internet use among middle and secondary school students*. The project co-financed from the resources of Gambling Problem Solving Fund being at disposal of Minister of Health. Main contractors: K. Warzecha, Ph.D. (manager of the project), E. Krzyżak-Szymańska, Ph.D., A. Wójcik, Ph.D., T. Żądło, Ph.D.

References

[1] Błachnio, A., Przepiórka, A., Rowiński, T.: Dysfunkcjonalne korzystanie z Internetu – przegląd badań. *Psychologia Społeczna* **4** (2014), 378-395.
[2] Bracha, Cz.: *Metoda reprezentacyjna w badaniu opinii publicznej i marketing*. Efekt, Warszawa, 1998.
[3] Davis, R. A., Gordon, L., Flett, D., Besser, A.: Validation of a new scale for measuring problematic Internet use: Implications for pre-employment screening. *CyberPsychology & Behavior* **5** (2002), 331-345.
[4] Grohol, J.: Too much time online: Internet addiction or healthy social interactions? *CyberPsychology & Behavior* **2**(1999), 395-401.
[5] Habrat, B. (red.): *Zaburzenia uprawiania hazardu i inne tak zwane nałogi behawioralne*. Instytut Psychiatrii i Neurologii, Warszawa, 2016.
[6] Hawi, N.S., Błachnio, A., Przepiórka, A.: Polish validation of the Internet Addiction Test. *Computers in Human Behavior* **48** (2015), 548-553.
[7] Izdebski, P., Kotyśko, M.: Problemowe korzystanie z nowych mediów. W: *Zaburzenia uprawiania hazardu i inne tak zwane nałogi behawioralne* (Habrat, B., ed.). Instytut Psychiatrii i Neurologii, Warszawa, 2016, 219-304.
[8] Jarczyńska, J.: Problematyczne używanie Internetu przez młodzież i młodych dorosłych – przegląd narzędzi do przesiewowej oceny tego zjawiska. *Przegląd Pedagogiczny* **1** (2015), 119-136.
[9] Jarczyńska, J.: *Zaangażowanie w hazard młodzieży gimnazjalnej i ponadgimnazjalnej. Diagnoza zjawiska i jego wybranych uwarunkowań*. Wydawnictwo Uniwersytetu Kazimierza Wielkiego, Bydgoszcz, 2016.
[10] Kończak, G.: *Testy permutacyjne. Teoria i zastosowania.* University of Economics in Katowice, Katowice, 2016.
[11] Makaruk, K., Wójcik, Sz.: Nadużywanie Internetu przez młodzież. Wyniki badania EU NET ADB. *Dziecko krzywdzone. Teoria, badania, praktyka* **1** (2013), 35-48.
[12] Morahan-Martin, J., and Schumacher, P.: Incidence and correlates of pathological Internet use among college students. *Computers in Human Behavior* **16** (2000), 13-29.
[13] Pawlikowski, M., Nader, IW., Burge,r C., Stieger, S., Brand, M.: Pathological Internet use – It is a multidimensional and not a unidimensional construct. *Addiction Research and Theory* **22** (2014), 166-175.
[14] Pawłowska, B., Potembska, E.: Objawy zagrożenia i uzależnienia od Internetu mierzonego Kwestionariuszem do Badania Uzależnienia od Internetu, autorstwa Pawłowskiej i Potembskiej u młodzieży polskiej w wieku od 13 do 24 lat. *Current Problems of Psychiatry* **12** (2011), 439-442.
[15] Pesarin, F., Salmaso, L.: *Permutation Tests For Complex Data*. *Theory, Applications and Software*. John Wiley & Sons, Chichester, 2010.
[16] Särndal, C.E., Lundström, S.: *Estimation in Surveys with Nonresponse*. Wiley, New York, 2005.
[17] Warzecha, K., Wójcik, A.: Analysis of the availability of Internet access for young people in schools of Silesia Province based on spatial statistics, In: *Proceedings of 33rd International Conference Mathematical Methods in Economics* (Martinčík D., Ircingová J., and Janeček P., eds.). University of West Bohemia, Plzeň, 2015, 912-918.
[18] Young, K.: Clinical assessment of Internet – addicted clients. In: *Internet addiction: A handbook and guide to evaluation and treatment* (Young K., de Abreu C. N., eds.). John Wiley & Sons, Inc., New York, 2011.

The level of implementation of Europe 2020 Strategy headline areas in European union countries

Andrzej Wójcik[1], Katarzyna Warzecha[2]

Abstract. Europe 2020 is a programme, as a result of the economic crisis from 2008-2009, makes clear that there is a necessity of reforms employment and setting the EU development priorities in order to avoid similar crisis in the future. The main objective of conducted analyses is examination of the progress of indicators in Europe 2020 Strategy headline areas (meaning the progress of headline indicators in five areas: employment, research and development (R+D), education, climate change and energy, poverty and social exclusion). To examine similarities between the EU countries as regards the level of Europe 2020 Strategy implementation Hellwig's development pattern method was used. The headline indicators of Europe 2020 Strategy proposed by Eurostat were the basis of the evaluation. The calculations were made with Excel programme.

Keywords: Europe 2020 Strategy, Hellwig's development pattern, indicators of Europe 2020 Strategy headline areas

JEL Classification: C10, Q01, R11
AMS Classification: 62F07, 62H86

1 Introduction

"Europe 2020– A strategy for smart, sustainable and inclusive growth" is a programme of socio- economic development of European Union countries for years 2010-2020 [3]. This strategy is to comply with long term challenges facing Europe which are connected with globalization, ageing of population or the rising need of rational use of the resources. The main aim of Europe 2020 Strategy is the economic growth and balancing of this process [14].

1.1 Objective, subject and research method

The present study is to examine the progress of indicators in Europe 2020 Strategy headline areas (meaning the progress of headline indicators in five areas: employment, research and development (R+D), education, climate change and energy, poverty and social exclusion (Table 1)). The described Europe 2020 Strategy is based on three mainstays connected with particular objectives that are set to be realized in 2020 (the objectives were presented in Table 1 for EU-28 as a whole and for Poland individually) [3]:

- sustainable growth (defined by four indicators)- which is connected with the development of efficient economy using the resources and protecting the environment thanks to low- emission management;
- inclusive growth (defined by two indicators)- which is connected with the development that is conductive to social integration with a special emphasis on creating new workplaces and reducing poverty, as well as with striving for professional activation of the largest possible number of persons;
- smart growth (defined by two indicators)- which is connected with economy development based on the knowledge and innovations thanks to investments in educational quality, scientific research and innovations.

Europe 2020 Strategy is a social vision of market economy for 21st century Europe, the economy with high level of employment and territorial cohesion [5]. The target values of Europe 2020 Strategy were defined on global level for the whole European Union (presented in Table 1) and for particular member countries as well. EU countries are diverse as regards society and economy. Therefore their objectives should be adjusted to their specificity and problem areas.

To evaluate the implementation of Europe 2020 Strategy by particular EU countries the taxonomic measure of Z. Hellwig was used (the description of measure: [5, 6, 7, 11, 12, 13]. Of the many methods of multidimensional comparative analysis [2, 4, 8, 10], this method was chosen because, it allows for the arrangement of research objects (EU countries) as regards the researched phenomenon. The level of Europe 2020

[1] University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, 50, 1 Maja Street, 40-287 Katowice, Poland, e-mail:andrzej.wojcik@ue.katowice.pl
[2] University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, 50, 1 Maja Street, 40-287 Katowice, Poland, e-mail:katarzyna.warzecha@ue.katowice.pl

Strategy implementation was characterized with the use of key variables (indicators) created to monitor the strategy (variables in division for stimulants and destimulants are presented in Table 1).

Strategy objectives	Indicators	EU objective as a whole	Polish objective
Objectives connected with sustainable growth			
1.Increase of investments in research and development sector	Expenditure on R+D (in % GDP) - S	3% GDP	1.7% GDP
2.Reduction of greenhouse gas emissions (CO_2)	Greenhouse gas emission (dynamics indices 1990=100)-D	20%	14%
3.Increase of the share of renewable energy sources in the whole energy consumption	The share of renewable energy sources in gross final energy consumption in % - S	20%	15.5%
4.Increase of energy efficiency	Primary energy consumption in Mtoe per 10000 people- D	20%	14%
Inclusive growth – connected with social inclusion			
5.Increase of employment rate of the population aged 20-64	Indicator of employment of population aged 20-64 in % - S	75%	at least 71%
6.Decrease of rate of people at risk of poverty and social exclusion in EU as a whole.	Indicator of rate of people at risk of poverty and social exclusion in % - D	by 20 million people	by 1.5 million people
Objectives connected with smart growth			
7. Reduction of percentage of early school leavers	Share of young people (aged 18-24) not continuing education in % of population in the same age in general-D	up to 10%	up to 4.5%
8. Increase of rate of people aged 30-34 with tertiary educational attainment	Share of people aged 30-34 with tertiary educational attainment in % of population in the same age in general-S	40%	up to 45%

S- stimulant, D – destimulant

Table 1. Objectives and headline indicators of Europe 2020 Strategy

Destimulants were changed into stimulants according to the pattern ($D \to S = \max x_{ij} - x_{ij}$) and then normalized (because the diagnostic variables in the research are given in various units of measure they cannot be submitted directly to aggregation) according to pattern 1 with the use of a zero unitarization method (for stimulants) [1]:

$$z_{ij} = \frac{x_{ij} - \min\{x_{ij}\}}{\max\{x_{ij}\} - \min\{x_{ij}\}} \tag{1}$$

where: min x_{ij}- minimum of variable x_i; max x_{ij} – maximum of variable x_i.

Having normalized variables, it is possible to define the distance from the development pattern P_0 between particular objects. The development pattern in this case is abstract point P_0 with standardized coordinates

$$P_0 = [z_{01}, z_{02}, ..., z_{0k}] \tag{2}$$

The coordinates of point P_0 are defined with the use of pattern 4 because all of the variables are stimulants.

$$z_{0j} = \max\{z_{ij}\} \tag{3}$$

Further on, the pattern 4 is used to calculate the Euclidean distance:

$$d_{i0} = \sqrt{\sum_{j=1}^{k} (z_{ij} - z_{0j})^2} \qquad (i = 1,2,...,m) \quad (j = 1,2,...,k) \tag{4}$$

where: k, m – are accordingly number of the variables and number of the objects,

z_{ij} – standardized values of variable j for object i; z_{0j} – standardized pattern value for variable j.

Subsequently, the Hellwig's taxonomic measure of development (Z_i) is being calculated, in other words, the synthetic measure of development which takes values within the range [0,1] (the exception is the situation when the development of particular object is definitely weaker than the development of other objects- and then it may take negative value [9]. The higher the values of this measure mean more favorable situation of particular object (country). The synthetic measure values mean higher level of Europe 2020 Strategy objectives implementation in particular countries.

$$Z_i = 1 - \frac{d_{i0}}{d_0} \tag{5}$$

where: $d_0 = \bar{d}_0 + 2S_0$; $\bar{d}_0 = \frac{1}{m}\sum_{i=1}^{m} d_{i0}$; $S_0 = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(d_{i0} - \bar{d}_0\right)^2}$

On the basis of Hellwig's taxonomic measure of development (Z_i) the division of researched countries into homogenous classes may be proposed (i.e. the classes with similar level of Europe 2020 Strategy objectives implementation). The researched countries were divided into four typological groups using the three- means method. The set of objects is divided into two subsets: first one contains the objects which are matched with measure values higher than the mean value, second one contains all of the remaining objects. Further on, indirect means will be calculated ([7], pp. 126-127):

Class I – high level of implementation of Europe 2020 Strategy objectives when $z_i > \overline{z_{1i}}$

Class II- moderate level of implementation of Europe 2020 Strategy objectives when $\overline{z_i} < z_i \leq \overline{z_{1i}}$

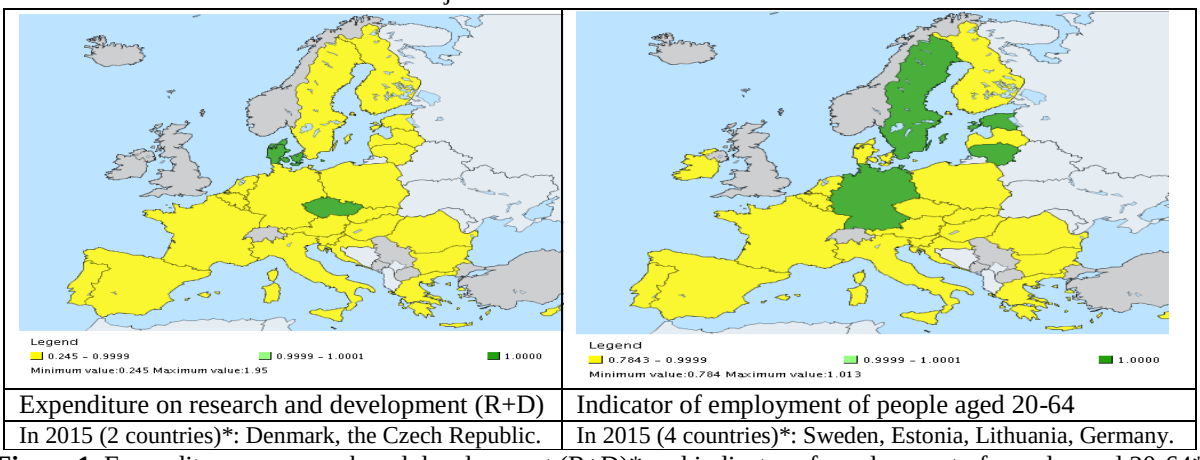
Class III- low level of implementation of Europe 2020 Strategy objectives when $\overline{z_{2i}} < z_i \leq \overline{z_i}$

Class IV- very low level of implementation of Europe 2020 Strategy objectives when $z_i \leq \overline{z_{2i}}$

where: $\overline{z_i}$ - mean taken from development measure value i , $\overline{z_{1i}}$, $\overline{z_{2i}}$ - indirect means taken from development measure value.

2 The implementation of Europe 2020 Strategy objectives in EU countries

From the data presented in Fig. 1. it results that in 2015 only two countries fulfilled the Europe 2020 Strategy objective concerning the increase of expenditure on R+D and they were: Denmark and the Czech Republic. While the Europe 2020 Strategy objective concerning the increase of employment rate in group of people aged 20-64 was fulfilled in 2015 by four countries: Sweden, Estonia, Lithuania and Germany. For comparison, in 2010 none of the countries fulfilled the objective.



Expenditure on research and development (R+D)	Indicator of employment of people aged 20-64
In 2015 (2 countries)*: Denmark, the Czech Republic.	In 2015 (4 countries)*: Sweden, Estonia, Lithuania, Germany.

Figure 1 Expenditure on research and development (R+D)* and indicator of employment of people aged 20-64*
* - countries marked in dark color fulfilled the Europe 2020 Strategy objective in particular year

Between 2010-2014 favorable changes were observed when it comes to climate change and energy objectives in Europe 2020 Strategy (Fig.2.). The positive changes concerning the use of renewable sources of energy were noted in all of the member countries. As it is shown in Fig.2., the Europe 2020 Strategy objective concerning the increase of renewable energy share in gross final energy consumption was fulfilled in 2014 by 9 EU countries, in comparison to 2010 when only 1 country achieved the goal- Croatia. In 2014 in Poland, the share of renewable energy in gross final energy consumption equaled to 11.45% and it was higher by 2 percentage points in comparison to 2010. Therefore, to achieve target value there are 4.05 percentage points missing. In Fig.2. countries that fulfilled the Europe 2020 Strategy objective concerning greenhouse gas emissions were presented (1990=100) - in 2014 seventeen of EU countries and in 2010 ten countries fulfilled the objective. Then, in majority of the EU countries the decrease of greenhouse gas emissions was noticed. The highest decrease in 2014 in comparison to 1990 was noticed in Lithuania (by about 60%), Latvia (by about 56%) and in Romania (by about 57%). In five member countries we may observe an alarming phenomenon, namely the increase of greenhouse gas emissions- in Cyprus (by 43.11%), Malta (by 50.88%), Spain (by 17.54%), Portugal (by 8.82%) and in Ireland (by 5.69%). In 2015 in EU, final energy consumption (indicated in tonnes of oil equivalent – TOE) equaled to 1529.6 million TOE (by 127.1 TOE less than in 2010), and in Poland 90 million TOE (by 5.7 million TOE less than in 2010). As it is shown in Fig.3. the Europe 2020 Strategy objective concerning the

increase of energy efficiency (i.e. the decrease of primary energy consumption) was fulfilled in 2015 by eleven EU countries.

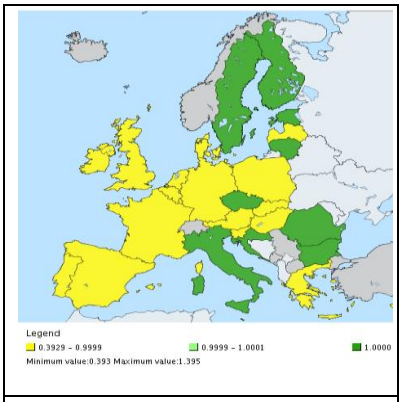
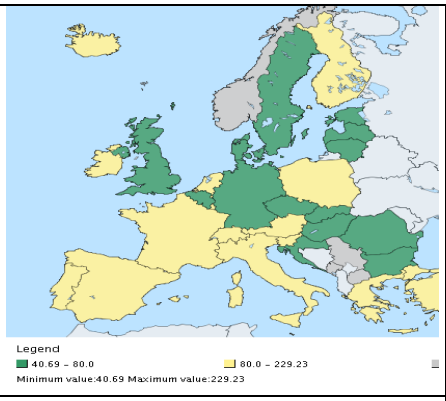
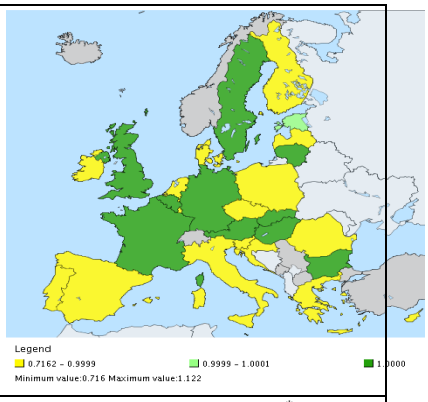
		
Share of renewable energy sources in gross final energy consumption	Greenhouse gas emissions (1990=100)	Primary energy consumption*
In 2014** (9 countries): Finland, Sweden, Estonia, Lithuania, the Czech Republic, Romania, Bulgaria, Croatia, Italy.	In 2014** (14 countries): Bulgaria, Romania, Hungary, Slovakia, the Czech Republic, Germany, United Kingdom, Lithuania, Latvia, Estonia, Denmark, Croatia, Sweden, Belgium; additionally 3 countries fulfilled only the national objective (Poland, Slovenia, Italy) .	In 2015 (11 countries): Sweden, Belgium, Bulgaria, Malta, United Kingdom, Germany, France, the Czech Republic, Ireland, Cyprus, the Netherlands.

Figure 2 Share of renewable energy sources in gross final energy consumption*, greenhouse gas emissions (1990=100)* and primary energy consumption* (** - no data available from 2015)
* - countries marked in dark color fulfilled the Europe 2020 Strategy objective in particular year

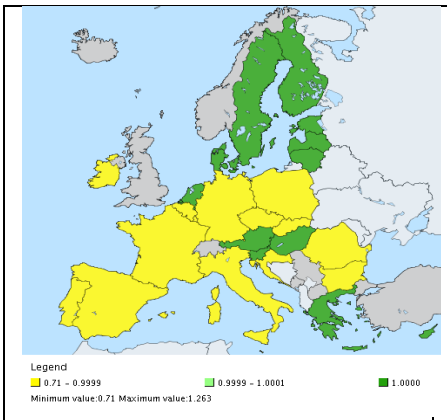
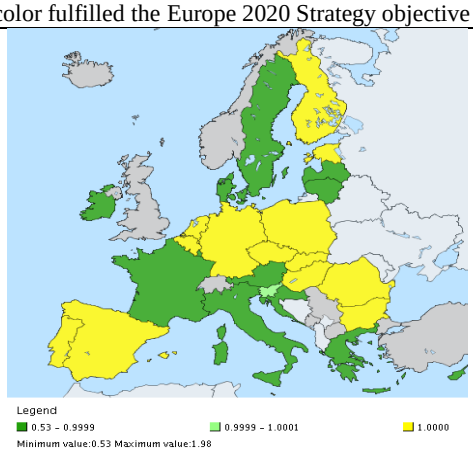
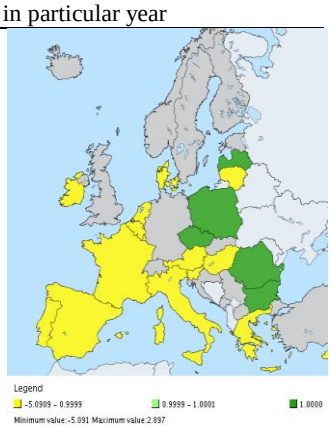
		
Share of people aged 30-34 with tertiary educational attainment in % of population in the same age in general*	Share of young people (aged 18-24) not continuing education in % of population in the same age in general*	The risk of poverty or social exclusion*
In 2015 (12 countries): Finland, Sweden, Denmark, the Netherlands, Estonia, Slovenia, Austria, Lithuania, Latvia, Greece, Cyprus, Hungary.	In 2015 (13 countries): Latvia, Lithuania, Luxembourg, Austria, Slovenia, Sweden, Denmark, Greece, Cyprus, Croatia, Italy, Ireland, France.	In 2015 (5 countries): Bulgaria, Romania, the Czech Republic, Poland, Latvia.

Figure 3 Percentage of people aged 30-34 with tertiary education* and the percentage of young people aged 18-24 with secondary education at most who did not continue the education* and the risk of poverty or social exclusion*
* - countries marked in dark color fulfilled the Europe 2020 Strategy objective in particular year

In European Union countries the share of people aged 30-34 with tertiary education equaled in 2015 to 39% and it was higher by 5 percentage points in comparison to 2010 (34%). According to Europe 2020 Strategy, this indicator should be equal to 40% by 2020. From the data presented in Fig.3. it results that in 2010 only five EU countries (Finland, Sweden, Denmark, the Netherlands, Estonia) and in 2015 twelve EU countries (Finland, Sweden, Denmark, the Netherlands, Estonia, Slovenia, Austria, Lithuania, Latvia, Greece, Cyprus, Hungary) fulfilled the Europe 2020 Strategy objective concerning the increase of percentage of people aged 30-34 with tertiary education. The lowest percentage of people with tertiary education was noted in 2015 and in 2010 in Italy (25.3% and 19.9%), Romania (25.6%; 18.3%), Malta (27.8%; 22.1%) and Slovenia (28.4%; 22.1%). As opposed to the countries where the university graduates were the majority in researched age group in 2015 and 2010. And they were: Lithuania (57.6%, 43.8%), Luxembourg (52.3%; 43.8%), Cyprus (54.6%; 45.3%), Ireland (52.3%; 50.1%) and Sweden (50.2%; 45.3%). Poland with the result 43.4% in 2015 was slightly higher than EU

mean result. The objective defined for our country as regards the Europe 2020 Strategy is 5.9 percentage point higher than the objective for the EU as a whole, by 2020 at least 45% of Poles should achieve tertiary education. Young people aged 15-16 often face the choice between further education, training or searching for a job. Obligatory education in full time lasts on average for 9-10 years in majority of European Union member countries (equivalent of Polish middle school). From the data presented in Fig.3. it results that in 2010 only eight EU countries (the Czech Republic, Latvia, Lithuania, Luxembourg, Austria, Slovenia, Slovakia, Sweden) and in 2015 thirteen EU countries (Latvia, Lithuania, Luxembourg, Austria, Slovenia, Sweden, Denmark, Greece, Cyprus, Croatia, Italy, Ireland, France) fulfilled the Europe 2020 Strategy objective concerning the percentage of young people aged 18-24 with secondary education at most who did not continue education. As results from Eurostat data, the highest number of people aged 18-24 dropped school in 2015 in: Spain (20%), Malta (19.8%), Romania (19.1%), Italy (14.7%). Poland was a part of the group with the lowest level of this indicator (5.3%), there were also Slovenia (5.0%), Cyprus (5.2%) and Croatia (2.8%). However, the Europe 2020 Strategy objective concerning the reduction of people being at risk of poverty and social exclusion (Fig.3.) was fulfilled in 2015 only by five countries: Bulgaria, Romania, the Czech Republic, Poland and Latvia.

The level of implementation of Europe 2020 Strategy objectives is diverse. In 2015 none of the countries fulfilled all of the national targets. Only Sweden fulfilled 6 out of 8 objectives as for now and Latvia 5 out of 8. Four Europe 2020 Strategy objectives were fulfilled by Bulgaria, the Czech Republic, Denmark, Estonia and Latvia. Poland fulfilled 2 objectives as for now (the increase of renewable energy share in final energy consumption and the decrease, in the EU scale, of number of people at risk of poverty and social exclusion). Two countries: Spain and Portugal did not fulfill any of the objectives. The most of the national Europe 2020 Strategy targets were fulfilled by EU countries in the area of sustainable growth.

3 Similarities of EU countries as regards comparable level of implementation of Europe 2020 Strategy

	EU countries	2010 Z_i	Typological group	*EU countries*	2015 Z_i	Typological group
1	Sweden	0.697		Sweden	0.677	
2	Denmark	0.609		Denmark	0.645	
3	Finland	0.541	Class I	Austria	0.528	
4	Estonia	0.466		Finland	0.494	
5	Austria	0.466	high level	Estonia	0.444	
6	Germany	0.460		Lithuania	0.441	Class I
7	France	0.433		The United Kingdom	0.409	
8	The United Kingdom	0.409		Germany	0.399	
9	The Netherlands	0.400		Latvia	0.384	
10	Belgium	0.370	Class II	France	0.379	Class II
11	Slovenia	0.370		The Netherlands	0.378	moderate level
12	Latvia	0.342	moderate level	Slovenia	0.356	
13	Lithuania	0.321		**P o l a n d**	0.346	
14	Ireland	0.315		Belgium	0.345	
15	**P o l a n d**	0.288		Ireland	0.338	
16	Croatia	0.270		Hungary	0.300	Class III
17	Bulgaria	0.266		Croatia	0.289	
18	Slovakia	0.221	Class III	Slovakia	0.272	low level
19	Hungary	0.216		Portugal	0.269	
20	Cyprus	0.205	low level	Bulgaria	0.262	
21	Portugal	0.205		The Czech Republic	0.218	
22	Luxembourg	0.201		Greece	0.189	
23	Greece	0.183		Cyprus	0.168	Class IV
24	Romania	0.183		Luxembourg	0.163	
25	The Czech Republic	0.167		Italy	0.131	very low level
26	Spain	0.157	Class IV	Romania	0.113	
24	Italy	0.127	very low level	Spain	0.076	
28	Malta	-0.048		Malta	-0041	

$\overline{z_i}$ = 0.320 , $\overline{z_{1i}}$ = 0.438, $\overline{z_{2i}}$ = 0.185, S= 0.160 for 2015; $\overline{z_i}$ = 0.316 , $\overline{z_{1i}}$ = 0.453, $\overline{z_{2i}}$ = 0.197, S= 0.158 for 2010

Table 2 Values of Hellwig's synthetic measure in 2010, 2015 for EU and the division for typological groups

Hellwig's taxonomic measure of development was used in the research to evaluate the implementation of Europe 2020 Strategy by EU countries and on the basis of this measure the evaluation of similarities between countries in 2010 and 2015 was made. Analyzed countries were divided as well into 4 typological groups (on the basis of calculated synthetic measure) with different levels of Europe 2020 Strategy implementation (Table 2). From the data in Table 3 results that in 2010 as well as in 2015 as regards the implementation of Europe 2020 Strategy such countries as: Sweden, Denmark, Austria, Finland, Estonia, Germany (only in 2010) and Lithuania (only in 2015) were in the leaders group. At the end of the list and the same in the group with very low level of Europe 2020 Strategy implementation were such countries as: Greece (only in 2010), Cyprus (only in 2015), Italy, Luxembourg, Spain, Malta, the Czech Republic (only in 2010) and Romania. In 2015 in comparison to 2010 the increase of Hellwig's synthetic measure value was noticed in 16 countries, the most significant in Lithuania. The most significant decrease was noticed in such countries as: Spain, Romania and France. In the EU countries rank, Poland went up from position 15 in 2010 (meaning Class III- low level) to position 13 in 2015 (meaning Class II- moderate level), and the synthetic measure rose by 0,059. The obtained results indicate the insignificant changes in synthetic measure values for the comparable periods of time. In 2015 in comparison to 2010 there was an increase of an average level (and the diversity of value of synthetic measure), and there was a high comparability of the changes direction of synthetic measure values (correlation coefficient 0.951).

4 Conclusions

The conducted research indicates that the level of implementation of Europe 2020 Strategy objectives by EU countries is diverse. The multidimensional comparative analysis- Hellwig's development pattern method was used and it enabled the evaluation of Europe 2020 Strategy implementation from the point of view of all of the indicators cumulatively and the definition which countries belong to one of the four typological groups (describing the best, moderate, poor and very poor situation as regards the implementation of the Strategy). The leadership among EU countries as regards the implementation of objectives of Europe 2020 Strategy is held by: Sweden, Denmark, Austria, Finland, Estonia, Germany (only in 2010) and Lithuania (only in 2015). At the end of the list, meaning in the group with very low level of Europe 2020 objectives implementation were such countries as: Greece, Cyprus, Italy, Luxembourg, Spain, Malta, the Czech Republic (only in 2010) and Romania (only in 2015). Two countries: Spain and Portugal did not fulfill any of the analyzed objectives. In the EU countries rank Poland went up from position 16 in 2010 (meaning Class III- low level) to position 12 in 2015 (meaning Class II- moderate level), and the synthetic measure rose by 0,064. In 2015 in comparison to 2010 the increase of the value of Hellwig's synthetic measure was noticeable in 16 countries, the most significant in Lithuania. The increase of value of synthetic measure indicates the gradual progress of the strategy implementation. The positive effects of Europe 2020 Strategy objectives implementation are clearly visible in indicators concerning the climate change and energy and education. However, two indicators concerning the poverty and social inclusion are still problematic for the EU countries.

References

[1] Balcerzak A.P.: Europe 2020 Strategy and Structural Diversity Between Old and New Member States. Application of Zero-unitization Method for Dynamic Analysis in the Years 2004–2013, *Economics & Sociology*, Vol. 8, No. 2/2015, 190–210.

[2] Balcerzak, A.P. and Pietrzak, M.P.: Application of TOPSIS Method for Analysis of Sustainable Development in European Union Countries. In T. Loster & T. Pavelka (Eds.). *The 10th International Days of Statistics and Economics. Conference Proceedings*. September 8-10, Prague, 2016, pp. 82-92.

[3] Europe 2020- A strategy for smart, sustainable and inclusive growth", European Commission, Brussels, 3.3.2010, Com (2010) 2020. http://ec.europa.eu/eu2020/pdf/1_PL_ACT_part1_v1.pdf [accessed: 1st March 2017]

[4] Giri Narayan C. *Multivariate Statistical Analysis*, USA Marcel Dekker, 2004, pp. 435-526.

[5] Greta, M., Tomczak, E., Lewandowski, K.: Strategia Lizbońska i Europa 2020 wobec społeczeństwa informacyjnego i gospodarki opartej na wiedzy. W: *Gospodarka elektroniczna. Wyzwania rozwojowe*. Zeszyty Naukowe Uniwersytetu Szczecińskiego, Szczecin, 2012, 82-92.

[6] Kuc M., The Implementation of Synthetic Variable for Constructing the Standard of Living Measure In European Union Countries , *Oeconomia Copernicana*, Vol. 3, No. 3/2012, 5-19.

[7] Młodak, A.: *Analiza taksonomiczna w statystyce regionalnej*, Difin , Warszawa, 2006.

[8] Mościbrodzka M., The Use of Methods of Multidimensional Comparative Analysis in Evaluation of the Standard of Living of Poland 's Population in Comparison with Other Countries of the European Union , *Oeconomia Copernicana*, Vol. 5, No. 3/2014, 29-47.

[9] Nowak, E. *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE, Warszawa 1990.

[10] Rencher A. C. *Methods of Multivariate Analysis* , USA, John Wiley & Sons, 2002.

[11] *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*, (Zeliaś, A. red.). AE, Kraków, 2000.
[12] Warzecha, K.: Poziom życia ludności Polski i pozostałych krajów Unii Europejskiej – analiza taksonomiczna. W: *Gospodarka polska po 20 latach transformacji: osiągnięcia, problemy i wyzwania*, (Pongsy-Kania, S., red.)., Instytut Wiedzy i Innowacji, Warszawa, 2009, 19-32.
[13] Warzecha K., Wójcik A.: Analysis of the availability of Internet access for young people in schools of Silesia Province based on spatial statistics. In: *Proceedings of 33rd International Conference Mathematical Methods in Economics (*Martinčík, D., Ircingová, J., Janeček, P., eds.). University of West Bohemia, Plzeň, 2015, 912-918.
[14] Wskaźniki Strategii Europa 2020 (Indicators to support Europe 2020 Strategy), http://stat.gov.pl/cps/rde/xbcr/gus/POZ_Wskazniki_Europa2020.pdf [accessed: 15th March 2017]
[15] Eurostat http://ec.europa.eu/eurostat/data/database [accessed: 15th March 2017]

On measuring accuracy in claim frequency prediction

Tomasz Żądło[1], Alicja Wolny-Dominiak[2], Wojciech Gamrot[3]

Abstract. Mean squared error (MSE) is the mean of squared errors which have positively skewed distribution. It means that the mean (here: MSE) should not be used as the only measure to describe the prediction accuracy even in terms of the average. The new measure of prediction accuracy based on quantiles of absolute prediction errors and the estimation procedure applicable for any parametric model are proposed. The estimation procedure is based on the parametric bootstrap method where absolute prediction errors are randomly generated under the assumed model. An application to claim frequency prediction in short term property/casualty insurance is given. Based on a real automobile insurance portfolio we show that even if estimated values of root mean squared errors (RMSE) for different forecasts are similar, values of estimates of the proposed measures could be substantially different what means that using only RMSE may cause a huge loss of information on prediction accuracy.

Keywords: prediction accuracy, mean squared error, parametric bootstrap, claim count, generalized linear mixed model.

JEL classification: C21, C25, C53
AMS classification: 62J12, 62M20

1 Introduction

Prediction theory finds numerous successful applications in economic decision making. Properties of different predictors are compared using sample statistics which measure various aspects and dimensions of prediction quality such as goodness-of-fit, risk or accuracy. For example, to compare relevance of two models, one can use among others the coefficient of determination, residual variance, Bayesian Information Criterion or Akaike's Information Criterion (see e.g. [12]). To compare risks of alternative portfolios, variance, semivariance, value at risk or conditional value at risk are often used. In some applications such as estimating index numbers the whole distribution of sample statistics may be of interest (e.g. [4]). One of the most popular prediction accuracy measures is the mean squared error of prediction and its modifications such as heteroskedasicity adjusted mean squared error (e.g. [9]).

In the paper we propose the quantile measure of prediction accuracy and the estimation procedure for a parametric model with fixed and random effects with an application in in claim frequency prediction, which is important in classification ratemaking (see e.g. [7], [1] and [17]).

2 Quantile based prediction accuracy measure - background

Let us consider two sets of data. The set A is a set of available sample data and it consists of n realizations of the variable of interest Y_i denoted by y_i and observations of auxiliary variables $\mathbf{w}_i$. In the prediction set P only auxiliary information on $\mathbf{w}_k$ is available (where $k = 1, ..., m$) for which we would like to predict values of Y_k denoted by y_k. Let $M(\mathbf{W}, \boldsymbol{\theta})$ be a model describing the class of distributions to which the distribution of Y_i (where $i = 1, ..., n$) and Y_k (where $k = 1, ..., m$) belongs, where $\mathbf{W}$ is a matrix of observations of auxiliary variables and $\boldsymbol{\theta}$ is a vector of unknown parameters. Let $\hat{Y}_k$ be a predictor of a random variable Y_k, under a model $M(\mathbf{W}, \boldsymbol{\theta})$. Hence $U = \hat{Y}_k - Y_k$ is a prediction error and Root Mean Squared Error (RMSE) is defined as $RMSE(\hat{Y}_k) = \sqrt{E(\hat{Y}_k - Y_k)^2} = \sqrt{E(U^2)}$. Let us propose an alternative to the RMSE, the Quantile of Absolute Prediction Error measure (QAPE) defined as the pth quantile (denoted by $Q_p(.)$) of absolute prediction error:

$$QAPE(p) = Q_p(|U|) = Q_p(|\hat{Y}_k - Y_k|) \tag{1}$$

Hence, at least $p\%$ of observed absolute prediction errors is smaller or equal $QAPE(p)$ while at least $1 - p\%$ of them is greater or equal $QAPE(p)$. Quantiles reflect the relation between the magnitude of the error and the

[1]University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, 50, 1 Maja Street, Katowice, Poland, tomasz.zadlo@ue.katowice.pl
[2]University of Economics in Katowice, Faculty of Economics, Department of Statistical and Mathematical Methods in Economics, 50, 1 Maja Street, Katowice, Poland, woali@ue.katowice.pl
[3]University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, 50, 1 Maja Street, Katowice, Poland, wojciech.gamrot@ue.katowice.pl

probability of its realization. It means that using QAPE one is able to describe fully the distribution of prediction errors as opposed to using the average (reflected by RMSE).

To estimate (1) we propose to use the well-known parametric bootstrap technique (see e.g. [6]). The fully parametric bootstrap procedure is as follows:

- based on n observations of $\mathbf{w}_i$ and y_i we estimate $\boldsymbol{\theta}$ and we obtain a vector of estimates $\hat{\boldsymbol{\theta}}$
- then, we generate B realizations of data for both sets A and P under the model $M(\mathbf{W}, \hat{\boldsymbol{\theta}})$. It means that we obtain B bootstrap realizations of Y_i (where $i = 1, ..., n$) denoted by $y_i^{*(b)}$; and B bootstrap realizations of Y_k (where $k = 1, ..., m$) denoted by $y_k^{*(b)}$, where $b = 1, ..., B$
- in the b-th iteration (where $b = 1, ..., B$) using $y_i^{*(b)}$ we obtain the vector of estimates $\hat{\boldsymbol{\theta}}^{*(b)}$, and hence we are able to obtain values of predictors $\hat{Y}_k^{*(b)}$ (denoted by $\hat{y}_k^{*(b)}$) for the set P based on the model $M(\mathbf{W}, \hat{\boldsymbol{\theta}}^{*(b)})$, where $k = 1, ..., m$
- finally, for the k-th element of the set P, we compute parametric bootstrap estimator of (1) as the pth quantile of absolute prediction errors in the bootstrap distribution:

$$\widehat{QAPE}(p) = \hat{Q}_p(|U|) = Q_p(|\hat{y}_k^* - y_k^*|). \tag{2}$$

If in the last step, we replace (2) by $\widehat{RMSE}(\hat{Y}_k) = (B^{-1} \sum_{b=1}^{B} (\hat{y}_k^* - y_k^*)^2)^{0.5}$, we obtain parametric bootstrap estimator of Root Mean Squared Error.

Similar accuracy measures were proposed by [18], but under simpler i.e. linear model and under normality. Moreover, they have been used for survey sampling purposes to measure accuracy of predictors of subpopulation totals. Here they are used to measure accuracy of predictors of the unknown value of the variable of interest.

3 An application in non-life insurance

We investigate the important aspect of risks pricing in insurance company, which is a'priori ratemaking in short time casualty/property insurance. The ultimate goal of a'priori ratemaking is grouping policies in a heterogenous portfolio of policies into homogenous classes and calculate the pure premium π_i for every ith individual policy, $i = 1, ..., n$. The pure premium for every policy in the portfolio is defined as follows:

$$\pi_i = S_i \cdot F_i \tag{3}$$

where n is the number of policies in portfolio, S is the claim severity and F is the claim frequency (see [3]). The estimation and prediction of variables S and F in (3) is required. In practice the GLM technique for cross-sectional data is used (see [10]). Focusing on F, let us denote the claim frequency for the i-th policy by Y_i and risk factors influencing the response variable by $x_{i1}, ..., x_{ip}$. The classical Poisson regression is assumed with log-link function as a special case of the GLM with $p = 1$ (see [10]). The proper GLM model $M(\mathbf{x}, \boldsymbol{\beta})$ is as follows:

$$\begin{cases} Y_i \sim Pois(\lambda_i), i = 1, ..., n \\ \lambda_i = \exp(\mathbf{x}_i^{'}\boldsymbol{\beta}) \\ g(x) = \log(x) \end{cases} \tag{4}$$

where $\mathbf{x}_i$ denotes the i-th row of the design matrix $\mathbf{X}$ for the ith policyholder, $\boldsymbol{\beta}$ denotes the vector for fixed effects estimated directly from data and $g(\cdot)$ is the link function. However model (4) is based on the assumption of independence among the responses $Y_1, ..., Y_n$. For some rating variables this assumption is clearly violated (see e.g. [2]), and in particular for the spatial variable . The solution is to incorporate the random effect into the model. We propose this approach in modelling the claims frequency applying GLMM Poisson (see [11]). The suitable model $M(\mathbf{x}, \mathbf{z}, \boldsymbol{\beta}, \sigma_v^2)$ is of the form:

$$\begin{cases} Y_i|\boldsymbol{v} \sim Pois(\lambda_i) \\ \lambda_i = \exp(\mathbf{x}_i^{'}\boldsymbol{\beta} + \mathbf{z}_i^{'}\boldsymbol{v}) \\ g(x) = \log(x) \\ v \sim N(0, \boldsymbol{I}\sigma_v^2) \end{cases} \tag{5}$$

where $\mathbf{z}_i$ is the spatial rating variable for the i-th insured and $\boldsymbol{v} = (v_1, ..., v_q)$ is the vector of random effects with the variance σ_v^2. To estimate parameters and to predict the number of claims $\hat{Y}_k$ we apply Penalized Quasi-Likelihood (see [5]). As the claims frequency in model (5) is a discrete variable and at the same time the model contains random effects, we adapt QAPE to obtain detailed information on prediction error distribution.

Based on the real automobile insurance portfolio taken from [8], we estimate the QAPE and classic RMSE using parametric bootstrap. The rating variables and categories of each risk factor in the model (5) are as follows: X_1 - Kilometres: an ordered factor representing kilomoters per year with five levels: less than 1000, 1000-15000, 15000-20000, 20000-25000, higher than 25000; X_2 - Bonus (the number of years, plus one, since last claim); X_3 - Make (common car models) and Z - Zone, a factor representing geographical area with the following levels 1: Stockholm, Goteborg, Malmo with surroundings, 2: Other large cities with surroundings, 3: Smaller cities with surroundings in southern Sweden, 4: Rural areas in southern Sweden, 5: Smaller cities with surroundings in northern Sweden, 6: Rural areas in northern Sweden, 7: Gotland. The details description is taken from the R package faraway. Underwriting and claims information are aggregated by categories, i.e. each combination of categories of risk factors corresponding to the number of claims and the number of insured (the exposure). The set A is the set of aggregated policies registered in the portfolio while the set P is the set of all combinations of categories unregistered yet.

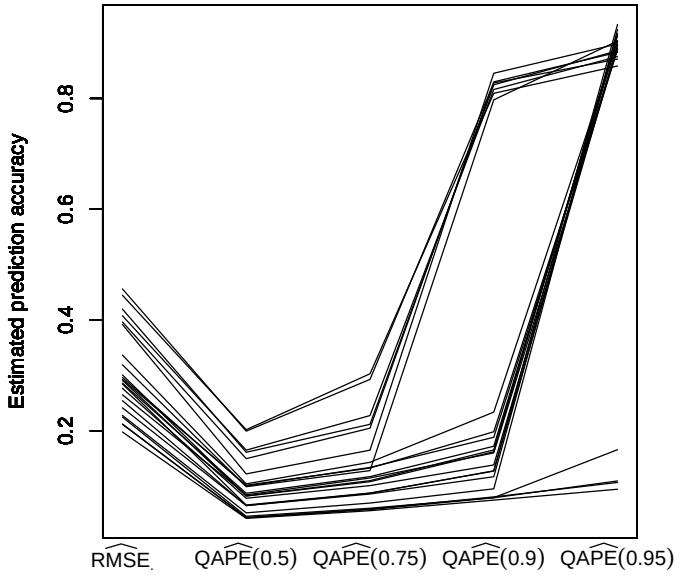


Figure 1 Parametric bootstrap estimates of RMSE's and QAPE's for all elements of P

To obtain the estimator of (1), we implement the parametric bootstrap procedure as in section 2. Moreover in the study we examine the behavior of the QAPE compared to RMSE estimated through the bootstrap procedure. The number of generated vectors of the study variable values based on the model (5) was 1000. The bootstrap model $M(\mathbf{x}, \mathbf{z}, \hat{\boldsymbol{\theta}}^*)$ referred to the model (5) is given by: $Y_i^*|\boldsymbol{v}^* \sim Pois(\exp(\mathbf{x}_i^{'}\hat{\boldsymbol{\beta}} + \boldsymbol{z}_i^{'}\boldsymbol{v}^*))$, and for the set P: $Y_k^*|\boldsymbol{v}^* \sim Pois(\exp(\boldsymbol{x}_k^{'}\hat{\boldsymbol{\beta}} + \boldsymbol{z}_k^{'}\boldsymbol{v}^*))$, where $\boldsymbol{v}^* \sim N(0, \boldsymbol{I}\hat{\sigma}_v^2)$. Moreover, $\hat{Y}_k^* = \exp(\boldsymbol{x}_k^{'}\hat{\boldsymbol{\beta}}^* + \boldsymbol{z}_k^{'}\hat{\boldsymbol{v}}^*)$, where $\hat{\boldsymbol{\beta}}^*$ and $\hat{\boldsymbol{v}}^*$ are Penalized Quasi-Likelihood estimators of $\boldsymbol{\beta}^*$ and $\boldsymbol{v}^*$ based on realizations of Y_i^*. The vector $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\beta}}^{'}, \hat{\sigma}_v^2]^{'}$ is Penalized Quasi-Likelihood estimator of $\boldsymbol{\theta} = [\boldsymbol{\beta}^{'}, \sigma_v^2]^{'}$.

In Figure 1 we present parametric bootstrap estimated values of RMSE and corresponding parametric bootstrap estimated values of QAPE for all of the 23 elements of the prediction set P. It is shown, that the distribution of RMSE within the set P is unimodal with small variability, while the distributions of quantiles (especially those of higher order) appear to be multimodal with much higher variability. For majority of elements of the set P (19 out of 23) the values of quantiles of order 0.95 are more than four times higher than values of quantiles of order 0.9. On the other hand, for four remaining elements of P values of these quantiles differ negligibly. Moreover, for the 4 elements quantiles of order 0.95 are smaller than the respective values of RMSE. Meanwhile for all of these 23 elements of P values of RMSE are similar which suggest similar prediction accuracy. In our study it is a consequence of discreteness of the variable of interest under assumed model. Hence, the Figure 1 clearly shows that using only RMSE may cause a huge loss of information on prediction accuracy.

Moreover, in Figure 2 we present parametric bootstrap distributions of absolute prediction errors for 4 cases chosen arbitrarily from among 23 elements of P. Values of QAPE estimated from these distributions are presented in Figure 1. Horizontal dotted lines indicate parametric bootstrap estimators of RMSE values. It is shown, that the bootstrap distribution of absolute prediction error is multimodal with strong positive skew. In one of the

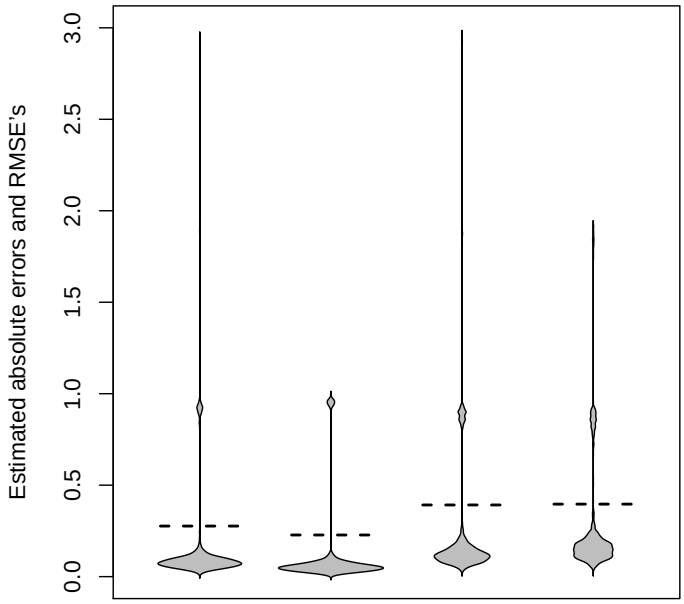


Figure 2 Bean plots of parametric bootstrap estimates of absolute prediction errors

presented cases estimated value of RMSE is approximately 12 times smaller than the maximum value of the bootstrap absolute prediction errors, what clearly shows the need to use other accuracy measures than RMSE.

4 Conclusions

Although the properties of the mean squared error are considered in the literature they are mainly discussed in terms of ex post prediction accuracy (see [15] and [16]) while we consider their properties in the case of ex ante prediction accuracy assessment. The mean squared error loss is sometimes replaced with asymmetric loss function (see [13]) or even with unknown loss function (see [14]) but the prediction accuracy is still studied in terms of the expected loss and the whole distribution of prediction errors is not of primary interest. Usually one would intuitively expect most prediction errors to take values relatively close to the root mean square error. Presented examples show, that this perception may be invalid. It may happen, that true errors are in most cases very distant from RMSE, or that realization of the true error close to RMSE is in fact impossible. In such cases the RMSE is seriously misleading and should not be used to measure the prediction accuracy. Instead, one should characterize the distribution of predictors using multiple quantiles of the absolute error.

References

[1] Antonio, K and Beirlant J.: Risk Classification in Nonlife Insurance, *Encyclopedia of Quantitative Risk Analysis and Assessment*, 2006.

[2] Antonio, K. and Beirlant J.: Actuarial statistics with generalized linear mixed models, *Insurance: Mathematics and Economics* **40** (2007), 58–76.

[3] Antonio, K and Valdez, EA.: Statistical concepts of a priori and a posteriori risk classification in insurance, *AStA Advances in Statistical Analysis*, **96** (2012), 187–224.

[4] Białek, J.: Proposition of a General Formula for Price Indices, *Communications in Statistics - Theory and Methods*, **41** (2012), 943–952.

[5] Breslow, N. E. and Clayton, D.G.: Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88** (1993), 9–25.

[6] Chatterjee, A. L. P., and Li, H.: Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models, *Annals of Statistics* **36** (2008), 1221–1245.

[7] Denuit, M and Maréchal, X and Pitrebois, S and Walhin, J-F.: *Actuarial modelling of claim counts: Risk*

classification, credibility and bonus-malus systems. Wiley, Chichester, 2007.

[8] Hallin, M., and Ingenbleek, J.-F.: The Swedish automobile portfolio in 1977. A statistical study, *Scandinavian Actuarial Journal* **88** (1983), 49–64.

[9] Małecka M.: Prognozowanie zmienności indeksów giełdowych przy wykorzystaniu modelu klasy GARCH. *Ekonomista* **6** (2011), 843–860.

[10] McCullagh, P. and Nelder, J.A.: *Generalized Linear Models. Second Edition*. Chapman & Hall, London, 1989.

[11] McCulloch, Ch. E., Searle, S. R.: *Generalized, linear, and mixed models*. Wiley, New York, 2006.

[12] Papież, M., Śmiech, S.: Causality in mean and variance between returns of crude oil and metal prices, agricultural prices and financial market prices In: *Proceedings of 30th international conference "Mathematical Methods in Economics 2012"* (Ramik, J. and Stavarek, D. eds.): Silesian University in Opava, School of Business Administration in Karvina. Karvina, 2009, 675–680.

[13] Patton, A. J., and Timmermann, A.: Properties of optimal forecasts under asymmetric loss and nonlinearity, *Journal of Econometrics* (2007), 884-918.

[14] Patton, A. J., and Timmermann, A.: Testing forecast optimality under unknown loss, *Journal of the American Statistical Association* (2007), 1172–1184.

[15] Swanson, D. A., and Coleman, Ch.: On MAPE-R as a measure of cross-sectional estimation and forecast accuracy, *Journal of Economic and Social Measurement* **32** (2007), 219-233.

[16] Swanson, D. A., Tayman, J., and Bryan, T. M.: MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecasts, *Journal of Population Research* **28** (2011), 225–243.

[17] Wolny-Dominiak, A.: Modeling of claim counts using data mining procedures in R CRAN. In: *Proceedings of 30th International Conference Mathematical Methods in Economics 2012* (J. Ramik,D. Stavarek eds.). Silesian University of Opava, School of Business Administration, Karvina, 2012, 980–985.

[18] Żądło, T.: On parametric bootstrap and alternatives of MSE. In: *Proceedings of 31st International Conference Mathematical Methods in Economics 2013* (Hana Vojáčková ed.), College of Polytechnics Jihlava, Jihlava, 2013, 1081–1086.

Multiple suppliers selection using the PROMETHEE V method

František Zapletal[1]

Abstract. A supplier selection problem is a common topic of many research studies. It is a typical problem of multi-criteria decision-making in practice. The aim of such studies is usually to choose the best (optimal) solution, to sort the suppliers into some classes or to find the non-dominated set of alternatives. This paper is devoted to the problem where a decision-maker (company) chooses a portfolio of suppliers to cover all his/her supply needs. It is shown here that it is insufficient to take the ranking of suppliers and to select the first few in order. The problem is solved by the PROMETHEE V method, which is a unique combination of the outranking method of multi-attribute decision-making and mathematical optimisation (first, the rankings of alternatives is found using the PROMETHEE II method, and then the optimal portfolio is selected based on that ranking using the integer programming). The problem is demonstrated on a numerical example at the end of the paper.

Keywords: PROMETHEE, integer programming, multi-criteria decision-making, outranking method.

JEL classification: C61
AMS classification: 90B50

1 Introduction

Almost each company providing services or selling some goods on a market has to face the problem of the supplier selection. Good and reliable suppliers help to better performance of a company. Therefore, it is natural to seek for an evaluation of the current or potential suppliers. The problem of the supplier selection is a typical example of multi-attribute decision making problem (MADM) where a discrete, finite and 'sufficiently' small set of alternatives is evaluated/selected based on a given set of criteria.

The author of [1] came with a thorough survey of scientific studies devoting to the supplier selection problem. That review shows a wide spectrum of methods which have been used for supplier selection as well as a huge number of published articles solving this problem. When taking into account only the multiple supplier selection, it is usual to use some combination of a MADM method and mathematical programming. The most popular is the Analytical Hierarchy/Network Process (AHP/ANP) either in the original version [2], [3] or its extension by different optimization models like integer programming [4], multi-objective integer programming [5], goal programming [6] and others. Approaches to the selection which use optimization methods directly also exist - for instance, Data Envelopment Analysis (DEA) [7] or the unique method of [8]. On the list of the used methods, we can find also some of so called outranking methods, which are also used in this paper, see, e.g. [9], or [10]. The aim of this paper is to demonstrate the applicability and suitability of using the PROMETHEE V method which is an extension of the PROMETHEE MADM method and to point out the corresponding advantages of this approach.

The paper is organized as follows. After this short introduction, a brief description of the family of PROMETHEE methods follows. The third section is devoted to the numerical example, which helps to demonstrate the applicability of the PROMETHEE V method for the supplier selection.

2 A family of the PROMETHEE methods

A group of MADM methods, which are based on pairwise comparison of alternatives using special preference (outranking) relations, is called the outranking methods. Those methods are partially compensatory (bad performance of the alternative with respect to one criteria can be to some extent compensated by excellent performances with respect to the remaining criteria). The main advantage of the outranking methods is the ability to deal with ordinal and more or less descriptive information on the alternatives under evaluation [11]. Outranking methods are often labeled as the European school of MADM because the origins of the most famous methods are in Europe (and most often in France). We may mention at least two main groups of the outranking methods - ELECTRE (established by Bernard Roy [12], [13]) and PROMETHEE (established by Brans and Vincke [14] and [15]). A brief description of other outranking methods can be found e.g. in [16]. If one searches the biggest citation databases, he/she finds out that these methods still grow in popularity.

[1]VŠB - Technical university of Ostrava, Sokolská 33, Ostrava, Czech Republic, frantisek.zapletal@vsb.cz

In this paper, the PROMETHEE method is used. Despite the fact that the ELECTRE methods are more complex, they require more input values, which are difficult or even impossible to interpret and I have to agree with [17] that the PROMETHEE is more tractable and more user-friendly for managers.

The PROMETHEE (Preference Ranking Organisation Method for Enrichment Evaluations) methods are based on the pair-wise comparisons of alternatives using the preference functions. And these preference functions $P_j : \mathbb{R} \to [0, 1]$ are defined on the basis of preference thresholds and they assign a preference level $P_j(a, b)$ to the difference in values of compared alternatives a and b regarding the j-th criterion. It is natural to require a non-decreasing shape of the preference function (greater differences in values imply the stronger preference). The most common shapes of preference functions can be found in [15]. Positive and negative flows of each alternative can be calculated using (1) and (2), see below. A positive flow can be understood as the index aggregating strengths of the alternative (i.e. areas in which the alternative has better performance when comparing with the other alternatives) and, similarly, a negative flow can be regarded as the index aggregating weaknesses of the alternative in comparison with other alternatives. Based on the corresponding equations, it can be seen that each better performance of an alternative within a pair-wise comparison increases the positive flow (and does not influence the negative flow) and vice versa. The PROMETHEE I method defines a preference relationship between alternatives using the formulas (3) and (4) below. In all other cases, a pair of alternatives is incomparable. Thus, the PROMETHEE I works with the partial ordering of alternatives. The PROMETHEE II enables a decision-maker to compare also pairs which are incomparable by the PROMETHEE I. It is done by net flows $\phi(a)$, which aggregate positive and negative flows to the only one index ($\phi(a) = \phi^+(a) - \phi^-(a)$), see (5). On the one hand, the complete ordering is set for ranking of alternatives. But on the other hand, a substantial part of information about alternatives' contradictions is lost here.

$$\phi^+(a) = \frac{1}{m-1} \sum_{b \neq a} \sum_{j=1}^{n} w_j \cdot P_j(a, b) \tag{1}$$

$$\phi^-(a) = \frac{1}{m-1} \sum_{b \neq a} \sum_{j=1}^{n} w_j \cdot P_j(b, a) \tag{2}$$

$$a \succ b \Leftrightarrow (\phi^+(a) \geq \phi^+(b) \wedge \phi^-(a) \leq \phi^-(b)) \tag{3}$$

$$a = b \Leftrightarrow (\phi^+(a) = \phi^+(b) \wedge \phi^-(a) = \phi^-(b)) \tag{4}$$

$$a \succ b \Leftrightarrow \phi(a) > \phi(b) \tag{5}$$

If the aim of decision-making is only to find a ranking of the alternatives or choose the best alternative, the task is completed by the PROMETHEE I or II, respectively. The situation becomes more complicated in the case that a decision-maker wants to select more than one alternative at once. An impetuos heuristic can tempt to easy-to-find solution driven only by the rankings of alternatives obtained before. For instance, if we want to select 3 cars within a car selection problem, it would mean to choose 3 pieces of the best alternative given by the rankings (in the case that the multiplicity of choice is allowed) or to select the alternatives which came first, second and third (if each alternative can be selected only once). In the supplier selection problem, it is natural to guarantee the uniqueness of alternatives in the selection. But, regardless the fact of the repeatability of alternatives in the selected set, the problem becomes more difficult to solve when some dependencies occur, i.e. when the choice of one option affects the suitability of choosing other variants. These dependencies can be caused by some additional constraints. Getting back to the car selection example, let us consider the budget constraint, which gives us the bound for the resulting set of alternatives. If we insist on the best one option (i.e. option with the highest net flow) as a part of the set of selected alternatives, it may lead to the situation that we will be forced to select two remaining cars from the bottom ranked variants.

As it has been already mentioned in the introduction, the use of mathematical programming to solve a MADM problem is not novel, see [4], [5], [6] or [8]. The PROMETHEE V method founded by [18] combines a discrete approach of MADM and optimization methods in a unique way. It extends the PROMETHEE II by restrictive constraints, which take into account that more than one alternative has to be chosen. Due to the fact that (unlike the approaches of [7] or [8]) variables represent directly selection (or non-selection) of options, a problem of integer programming is solved within the PROMETHEE V. Moreover, if each alternative can be chosen only once into the resulting set, the model is even more restricted and it leads to the 0-1 programming problem. A general form of a PROMETHEE V model is given by (6).

$$\begin{aligned} \max \quad & \boldsymbol{\phi}^T \boldsymbol{x} \\ \text{s.t.} \quad & \mathbf{A}\boldsymbol{x} \leq \mathbf{b} \\ & \boldsymbol{x} \in \mathbb{Z}, \end{aligned} \tag{6}$$

where $\boldsymbol{\phi} \in \mathbb{R}^{m\times 1}, \boldsymbol{\phi}^T = \big(\phi(a_1), \phi(a_2), \ldots, \phi(a_m)\big)$, i.e. vector of net flows from the PROMETHEE II. Note that the variables are integers ($\boldsymbol{x} \in \mathbb{Z}^{m\times 1}$) and they are interpreted as a number of the i-th alternative chosen. If it is not possible to select more than 1 'piece' of an alternative, the variables should be binary instead ($\boldsymbol{x} \in \{0, 1\}$). The constraints of the model with the coefficient matrix $\mathbf{A} \in \mathbb{R}^{r\times m}$ (r stands for a number of constraints) and right-hand-side coefficients $b \in \mathbb{R}^{r\times 1}$ delimit feasible combinations of the selected alternatives.

The PROMETHEE V method has been already used to solve some real economic applications, e.g. in engineering (rehabilitating of the water distribution network) by [19], in project prioritization [20] or in water resource planning [21].

In the survey paper by [1], some general requirements and often drawbacks for the method suitable for a supplier selection are given:
1. not so exhaustive time complexity to get the results;
2. good tractability;
3. possibility to involve uncertainty;
4. influence of a rank reversal.

The first two items on the list are satisfied. The first part of the procedure (PROMETHEE II) is extremely easy and fast to compute from the mathematical point of view. The most demanding is to deal with the optimization model. But, it is a model of linear programming where the existence of the optimal solution is guaranteed (if at least one feasible solution exists). Moreover, a structure of the linear programming model is easy to understand and well interpretable - there is no artificial measure (apart from net flows in the coefficients of the objective function) or the artificial variable. To some extent, it is also possible to include the uncertainty into the problem, because it can be expressed by a preference function's shape. To be able to involve more types of uncertainty, any further extension of the PROMETHEE method would have to be established. Unfortunately, the PROMETHEE methods suffer from the rank reversal (i.e. a possible inconsistency in the rankings when some small change in the problem's structure is made), see [22] to see more about this topic.

A problem of the PROMETHEE V (i.e. the problem of integer linear programming) can be solved by some algorithms for integer optimization, e.g. by cutting plane methods or the branch-and-bound method. Almost all optimization software allows for integer restriction for the variables, so the GAMS software or the simple Solver of MS Excel can be used. Moreover, one solver is implemented directly in the Visual PROMETHEE software, which is also used in this study.

3 A case study: Application of the PROMETHEE V in supplier selection

It is not the goal of the paper to evaluate or discuss suitable and non-suitable criteria, which should be taken into account. That is the reason why the existing structure of the model is used as a numerical example, in particular, the alternatives and their performances were taken from [8] and they are used as the input data for the PROMETHEE II analysis. The emphasis is put rather on a structure of the optimization model of the PROMETHEE V.

Fig. 1 shows the input values of the problem, which consists of 18 alternatives (suppliers) and 5 criteria for evaluation. Because the original study of [8] does not require the weights of criteria as the inputs for the analysis, they cannot be taken from there and they are set equal without loss of generality. The last task to be able to perform the PROMETHEE II analysis is to assign a preference function to each criteria. As it can be seen in Fig. 1, the linear shape and the V-shape are used. Both are piece-wise linear, but the former one uses an indifference threshold, i.e. the maximum positive value of difference in alternatives' performances which is negligible for a decision-maker. The values of the thresholds are subjective and they are set to values, which do not contradict any basic logic.

In line with the algorithm described in the previous section, the results of the PROMETHEE I and the PROMETHEE II ranking are found and shown in Tab. 1. The alternatives are ordered with respect to their net flows.

Now, let us assume that the company uses 10 different inputs for production. But none of the suppliers under evaluation is able to supply all inputs and equal amount of the inputs. Thus, the company wants to find such set of suppliers, which will supply the required inputs all together and with a minimum sum of the net flows. In Tab. 2, the amounts of 10 considered inputs (I1-I10) and the available amounts for all the suppliers are displayed. Then, let us suppose that the company needs various volumes of the inputs $(15, 60, 70, 120, 80, 70, 40, 100, 100, 100)$ units.

The results are displayed in Tab. 3. When no limit for the number of suppliers is given, the optimal portfolio

Rank	Supplier's code	ϕ	ϕ^+	ϕ^-	Rank	Supplier's code	ϕ	ϕ^+	ϕ^-
1	S15	0.4008	0.4373	0.0365	10	S1	-0.0167	0.1290	0.1457
2	S17	0.3953	0.4363	0.0410	11	S9	-0.0438	0.0869	0.1307
3	S10	0.2204	0.03429	0.1225	12	S16	-0.0671	0.0964	0.1635
4	S5	0.1546	0.2274	0.0728	13	S3	-0.0994	0.0686	0.1680
5	S8	0.0915	0.2045	0.1131	14	S2	-0.1399	0.0822	0.2221
6	S11	0.0794	0.1725	0.0931	15	S18	-0.1866	0.0494	0.2360
7	S6	0.0513	0.2333	0.1819	16	S4	-0.2879	0.0405	0.3285
8	S13	0.0498	0.1444	0.0946	17	S7	-0.3005	0.0363	0.3368
9	S12	0.0246	0.1392	0.1146	18	S14	-0.3258	0.0454	0.3712

Table 1 The results of the PROMETHEE II analysis

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
I1	10	0	0	20	0	0	0	0	0	0	0	0	30	0	0	0	0	0
I2	0	0	30	40	0	0	0	30	0	0	80	0	0	60	0	0	0	0
I3	0	0	0	50	0	0	20	0	0	0	0	0	50	0	0	0	0	0
I4	0	20	0	0	100	0	0	0	0	0	50	0	0	80	0	0	100	0
I5	0	0	0	0	40	30	0	0	40	0	0	0	50	0	50	0	30	0
I6	0	25	0	0	50	0	0	60	0	60	0	0	0	0	0	60	0	0
I7	0	0	0	0	0	0	0	50	0	60	0	50	0	0	50	0	0	0
I8	0	0	0	0	0	0	90	0	0	80	0	0	0	0	0	0	90	0
I9	0	0	0	0	0	0	0	0	0	120	0	0	0	0	80	0	0	70
I10	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	100

Table 2 Amounts of inputs which can be supplied

Max number of the suppliers	Optimum of the objective function	Optimal portfolio of the suppliers
without limit	0.9932	S4, S5, S6, S8, S10, S11, S12, S13, S15, S17, S18
10	0.9686	S4, S5, S6, S8, S10, S11, S13, S15, S17, S18
9	0.9173	S4, S5, S8, S10, S11, S13, S15, S17, S18
8	0.8379	S4, S5, S8, S10, S13, S15, S17, S18
7	0.4371	S4, S5, S8, S10, S13, S17, S18
6	0.0171	S5, S7, S10, S11, S13, S18

Table 3 The results of the PROMETHEE V

consists of 11 suppliers. But, it may be reasonable to put some restriction on that quantity (because of higher administrative costs or for other reasons). In Tab 3, the optimal solutions for maximum allowable number of suppliers in the portfolio from 6 to 10 are shown. If this limit is reduced to 5, no feasible solution exists. It is worth noting that the best two options of the PROMETHEE II solution (the suppliers S15 and S17) are not included when not more than 6 suppliers can be chosen. This points to the potential added value of the PROMETHEE V in selection. Constraints resulting from the demands on the optimal portfolio can cause that the portfolio differs from the rankings obtained by the PROMETHEE II method.

	Bertrand		Supply Variety	Quality	Distance	Delievery	Price index
	Unit		Pieces	% non-defect	km	% in time	%
	Cluster/Group						
	Preferences						
	Min/Max		max	max	min	max	min
	Weight		0,20	0,20	0,20	0,20	0,20
	Preference Fn.		Linear	V-shape	Linear	V-shape	V-shape
	Thresholds		absolute	absolute	absolute	absolute	absolute
	- Q: Indifference		5,000	n/a	30,0	n/a	n/a
	- P: Preference		20,000	5,00	1000,0	20,0	11,0
	- S: Gaussian		n/a	n/a	n/a	n/a	n/a
	Statistics						
	Minimum		1,000	91,89	238,0	80,0	80,0
	Maximum		53,000	100,00	1809,0	100,0	100,0
	Average		16,000	99,15	715,8	91,4	97,8
	Standard Dev.		13,812	1,98	396,1	4,9	6,3
	Evaluations						
☑	S1		2,000	100,00	249,0	90,0	100,0
☑	S2		13,000	99,79	643,0	80,0	100,0
☑	S3		3,000	100,00	714,0	90,0	100,0
☑	S4		3,000	100,00	1809,0	90,0	100,0
☑	S5		24,000	99,83	238,0	90,0	100,0
☑	S6		28,000	96,59	241,0	90,0	100,0
☑	S7		1,000	100,00	1404,0	85,0	100,0
☑	S8		24,000	100,00	984,0	97,0	100,0
☑	S9		11,000	99,91	641,0	90,0	100,0
☑	S10		53,000	97,54	588,0	100,0	100,0
☑	S11		10,000	99,95	241,0	95,0	100,0
☑	S12		7,000	99,85	567,0	98,0	100,0
☑	S13		19,000	99,97	567,0	90,0	100,0
☑	S14		12,000	91,89	967,0	90,0	100,0
☑	S15		33,000	99,99	635,0	95,0	80,0
☑	S16		2,000	100,00	795,0	95,0	100,0
☑	S17		34,000	99,99	689,0	95,0	80,0
☑	S18		9,000	99,36	913,0	85,0	100,0

Figure 1 Input data in Visual PROMETHEE SW [source: modified [8]]

4 Conclusions

The paper focused on the application of the PROMETHEE V method for a supplier selection problem. The study has shown that this method is very suitable in the case that more than one supplier should be chosen. From the mathematical point of view, the method is very simple to use and not so time-consuming. Moreover, this approach can be valued for its good tractability. On the other hand, it is not always easy (and sometimes it is even impossible) to define the set of constraints for the model. This paper is planned to be extended further in near future. It is challenging to establish a fuzzy extension of this approach, which has not been published yet and which may help to even better applicability of the PROMETHEE V in business practice.

Acknowledgements

This paper was supported by grant No. GA 16-01298S of the Czech Science Foundation, SGS project No.SP2017/125 and the Operational Programme Education for Competitiveness Project CZ.1.07/2.3.00/20.0296. The support is greatly acknowledged.

References

[1] Mukherjee, K.: Supplier selection criteria and methods: past, present and future, *International Journal of Operational Research* **27**, (2016), 356-373.

[2] Yao, F., and Hongli, L.: Information Systems Outsourcing Vendor Selection Based on Analytic Hierarchy Process, *IEEE*, (2007), 6250-6253.

[3] Hsu, C.-W., and Hu, A. H.: Applying hazardous substance management to supplier selection using analytic network process, *Journal of Cleaner Production*, **17**, (2009), 255-264.

[4] Yu, J.-R., and Tsai, C.-C.: A decision framework for supplier rating and purchase allocation: A case in the semiconductor industry, *Computers & Industrial Engineering* **55**, (2008), 634-646.

[5] Kokangul, A., and Susuz, Z.: Integrated analytical hierarch process and mathematical programming to supplier selection problem with quantity discount, *Applied Mathematical Modelling* **33**, (2009), 1417-1429.

[6] Erdem, A. S., and Gocen, E.: Development of a decision support system for supplier evaluation and order allocation, *Expert Systems with Applications* **39**, (2012), 4927-4937.

[7] Ha, S. H., and Krishnan, R.: A hybrid approach to supplier selection for the maintenance of a competitive supply chain, *Expert Systems with Applications* **34**, (2008), 1303-1311.

[8] Ng, W. L.: An efficient and simple model for multiple criteria supplier selection problem, *European Journal of Operational Research* **186**, (2008), 1059-1067.

[9] Boer, L. d., Wegen, L. w. d., and Telgen, J.: Outranking methods in support of supplier selection, *European Journal of Purchasing & Supply Management* **4**, (1998), 109-118.

[10] Sinaga, T.S., and Siregar, K.: Supplier Selection based on the Performance by using PROMETHEE Method. *In Proceedings of the 1st Annual Applied Science and Engineering Conference. IOP Conference series: Material Science and Engineering*, (2017).

[11] Kangas, A., Kangas, J., and Pykäläinen, J.: Outranking methods as tools in strategic natural resources planning, *Silva Fennica* **35**, (2001), 215-227.

[12] Roy, B.: ELECTRE III : un algorithme de classement fonde sur une representation floue des preferences en presence de criteres multiples, *Cahiers du CERO* **20**, (1978), 3-24.

[13] Roy, B., and Bertier, P.: La meethode ELECTRE II : une application au media-planning. *In M. Ross, editor, OR'72*, 291-302, 1973.

[14] Brans, J. P: *L'ingenierie de la decision: elaboration d'instruments d'aide a la decision. La methode PROMETHEE.* Presses de lUniversit Laval, 1982.

[15] Brans, J. P., and Vincke, P.: A preference ranking organisation method: The PROMETHEE method for MCDM, *Management Science*, 1985.

[16] Figueira, J., Greco, S., and Ehrgott, M.: *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer International Series, 2005.

[17] Geldermann, J., Spengler, T. S., and Rentz, O.: Fuzzy Outranking for Envi- ronmental Assessment - Case study: Iron and Steel Making Industry. *Fuzzy Sets and Systems* **115**, (2000), 45-65.

[18] Brans, J. P., Vincke, Ph., and Mareschal, B.: How to select and how to rank projects: The PROMETHEE method. *European Journal of Operational Research* **24**, (1986), 228-238.

[19] Fontana, M. E., and Morais, D. C. Using Promethee V to Select Alternatives so as to rehabilitate Water Supply Network with Detected Leaks, *Water Resource Management* **27**, (2013), 4021-4037.

[20] Mavrotas, G., Diakoulaki, D., and Caloghirou, Y.: Project prioritization under policy restrictions. A combination of MCDA with 01 programming, *European Journal of Operational Research* **171**, (2006), 296-308.

[21] Abu-Taleb, M., and Mareschal, B.: Water resources planning in the Middle East: Application of the PROMETHEE V multicriterion method, *European Journal of Operational Research* **81**, (1995), 500-511.

[22] Verly, C., and Smet, Y. d.: Some results about rank reversal instances in the PROMETHEE methods, *International Journal of Multicriteria Decision Making* **3**, (2013), 325-345.

The Impact of Statistical Standards on Input-Output Analysis

Jaroslav Zbranek[1], Jakub Fischer[2], Jaroslav Sixta[3]

Abstract. Symmetric input-output tables belong to the one of the most efficient tools for economic models. These models can be either static or dynamic and in all cases, they are based on the description of relationships between supplies and uses. Input-output tables were incorporated into national accounts in 1960s and since 1993 they belong to the key part of national accounts. This ensures the comparability with gross domestic product and other aggregates that are the subjects of researchers studies. This clear advantage is compensated by often and regular updates of national accounts figures due to the changes in statistical data sources or national accounts standards. These changes usually lead to non-negligible amendments of relations described by input-output tables. The recent standard SNA 2008 brought lots of conceptual changes that influenced input-output based technical coefficients. The aim of the paper is the illustration of dependency of input-output analysis on statistical methodology.

Keywords: input-output, technical coefficients, macroeconomic analysis, models.

JEL classification: C67, O11
AMS classification: 65C20

1 Introduction

Input-output tables (IOTs) provide valuable analytical tool that is usually used by the most qualified researchers. Using IOTs by academic sphere has a long tradition since W. Leontief presented his work in 1930s. Since then, the development in official statistics corresponds to the needs of researchers. With respect to the conservatism in academics, implementation of new statistical classifications and standards is not accepted without problems. It can be illustrated on the obsolete mark sector for columns of IOTs since they are labelled industries or products depending on the type. Sector currently refers to the institutional sector and sector accounts and that means something completely different. The family of IOTs currently cover symmetric input-output tables (SIOTs) and supply and use tables (SUTs). In modern economic statistics both SIOTs and SUTs are compiled under the common framework of national accounts and that ensures full consistency with national accounts macroaggregates, such as gross domestic product. Besides national figures, this has an impact on the regional IOT, as well. Since the models prepared on the basis of regional input-output tables (e.g. Šafr [8]) may provide different results.

In line with European practice, SIOTs are published every five years in currently valid standard ESA 2010. Such standard was put into force in September 2014 and it brought significant differences to national accounts Standards. Modern issues with globalisation or current state of economy cover activities such as research and development. One of the most important issues for the users of IOTs is that these data do not come directly from statistical surveys. The links between national and business accounting is a complex issue with many consequences. On the contrary, the complexity of SUTs and SIOTs ensures direct use in economic analyses.

The aim of the paper is the illustration of dependency of input-output analysis on statistical methodology. The impact of different standards is discussed on the case of simple static input-output analysis.

2 Methodology

Modern world brought economic events that had not been present or were not significant in the past. It means that statistical standards must react on the development of the society. In the field of macroeconomic statistics, new standard System of National Accounts 2008 (SNA 2008) was prepared. European modification ESA 2010 provided obligatory schemes of new IOTs and SUTs. The most important issues given by the updated statistical standards cover different approach to foreign trade and new assets. The issue of deflation was deeply discussed in Musil and Cihlář [4] and Kramulová et al. [2] and so we focus on the estimates at current prices only.

[1]University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, jaroslav.zbranek@vse.cz
[2]University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, fischerj@vse.cz
[3]University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3, Czech Republic, sixta@vse.cz

2.1 Strict ownership concept of foreign trade

The definition of foreign trade in macroeconomic statistics was significantly improved. The concept of economic ownerships influenced recording of processing, merchanting and re-exports. Processing was in line with previous ESA 1995 recorded differently to companies accounts. It reflected the process of production. In other words, it means that if a contractor producing goods on the material sent by the customer (who is the actual owner), the value of a material was imputed into the contractors intermediates for obtaining complete value of finished products. It resulted in the changes of exports and imports for the value of material used for further processing, details can be seen in Sixta et al. [5]. On the contrary, today only the service invoiced by the contractor is included. That may lead to paradox and incorrect conclusions about the production without intermediates. Even if this recording is closer to business accounting, the productivity analysis is more complicated.

The second point relates to the merchanting that express the value of re-sold goods abroad. The seller buys goods in a country *A* and sells with a margin in country *B*. The problem is that such operations are recorded on the export side only as a difference between sales and purchases. It means that purchasing goods for further re-sale is recorded a negative export and if the time difference between sales and purchases is significant, exports of some products can be negative. Obviously, this is a problem for input-output analysis and it causes interpretations difficulties as describes Sixta [6].

One of the most difficult problem to input-output analysis is caused by so-called re-exports. Re-exports represent goods imported and exported without any interaction with domestic economy. In national accounts, it is recorded in both exports and imports at the same value. It means that export of product can occur even if the product is not domestically produced. When the use of domestic output is computed that must be taken into account. It means that exports can originate from inventories instead of output.

2.2 New assets

SNA 2008 and ESA 2010 updated the definition of assets since new assets were introduced. The most important changes consist in the recording of expenditures on research and development, small tools and military assets. It means that these expenditures previously recorded in intermediate consumption belong to the gross fixed capital formation. These changes influenced technical coefficients and subsequently the results of the input-output analysis.

Expenditures on research and development represent more than 2% of gross domestic product in the Czech Republic. The recording of intermediate consumption of the product research and development (CZ-CPA 72) was changed. Output of such product is imputed as well as gross fixed capital formation. Such change is easily observable in SUTs/SIOTs.

Small tools represent the category of assets covering laptops, mobile phones, small devices that can be used in the production process more than one year. In ESA 1995 a threshold of 500 ECU at 1995 prices was given and currently no threshold exists. It means that companies intermediates is higher than in national accounts and vice versa in investments.

Military assets correspond to concept of defence services. Weapons are treated in the same way as other machinery and equipment since they provide a service, defence. The service is provided up the depreciation of the weapon or the weapon is destroyed. Since the weapons are purchased irregularly by the government, the recording may cause unpredictable development in the input-output analysis.

2.3 Principles of Input-Output Analysis

The simple static IOA is based on the use of static Leontiefs IOTs, which represents the easiest tool for making input-output analysis. The static Leontiefs IOTs based on inverse matrix as describes Hronová et al. [1] or Miller and Blair [3] shows following formula 1:

$$x = (I - A)^{-1} y, \tag{1}$$

where

x	output vector,
y	vector of final use,
$(I - A)^{-1}$	Leontief inverse.

The impact of the change in final use is measured by the change of output according to formula 2:

$$\Delta x = (I - A)^{-1} \Delta y, \tag{2}$$

In general, the simple static IOTs is used to analyse the impact on the national economy based on change of final use. This tool provides information on changed level of output, intermediate consumption, gross value added, compensation of employees and any other indicators. The most common use of simple statistic input-output analysis is the description of external investment shock.

3 Input-Output Analysis under a Different Methodology

In order to determine the impact of changes in statistical standards on IOA results, we performed a simple IOA assuming a one-time investment into public infrastructure (*CPA 42*) in the Czech Republic for *CZK 4 billion*. We conducted the IOA using an input-output model based on SIOT compiled according to ESA 1995 as well as from ESA 2010. This IOA was based on SIOT for the same assumptions for 1995, 2000, 2005 and 2010. For these years, the CZSO published SIOT reports according to both standards ESA 1995 and ESA 2010.

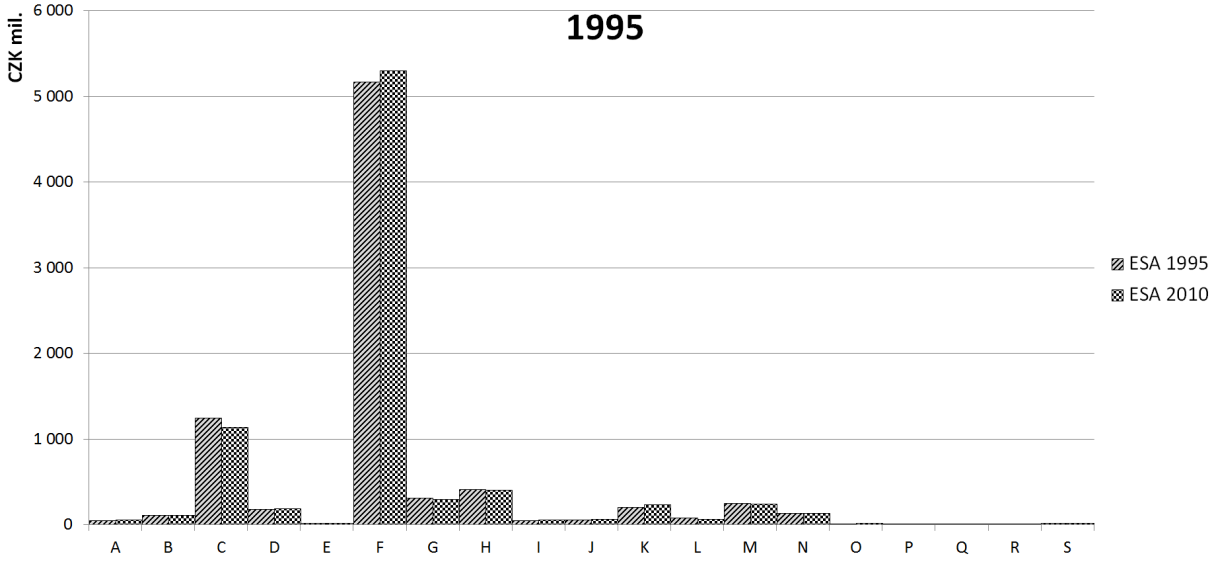


Figure 1 Comparison of results of input-output analysis based on SIOT according to ESA 1995 and ESA 2010 for 1995

The comparison of the IOA results for 1995 is shown in Figure 1[1]. It is clear from the graph that the IOA provides very similar results using SIOT according to ESA 1995 and ESA 2010. Minor differences are visible in the case of the production of construction products (*F*, $+2.6\%$) and products of the manufacturing industry (*C*, -8.6%). The total relative change of output between results based on ESA 1995 and ESA 2010 is $+0.3\%$ in 1995. In the case of the IOA for the year 2000, as shown in Figure 2[1], minor differences are observed in relation to the induced production of the products of the manufacturing industry (*C*; -8.3%). The total relative impact of the change of national account standards on the output in 2000 is -1.0%.

In the case of the IOA results for the 2005 (Figure 3[1]), smaller differences are seen in the case of the induced production of construction products (*F*, -3.0%) and professional, scientific and technical services (*M*, $+13.6\%$). The total relative change between results based on ESA 1995 and ESA 2010 for output in 2005 is -0.4%. The most significant differences in IOA's results are apparent from the 2010 data, as shown in Figure 4[1]. Significant differences can be observed in the case of induced production of construction products (*F*, 6.4%) and professional,

[1] **A** Products of agriculture, forestry and fishing; **B** Mining and quarrying; **C** Manufactured products; **D** Electricity, gas, steam and air conditioning; **E** Water supply; sewerage, waste management and remediation services; **F** Constructions and construction works; **G** Wholesale and retail trade services; repair services of motor vehiclesand motorcycles; **H** Transportation and storage services; **I** Accommodation and food services; **J** Information and communication services; **K** Financial and insurance services; **L** Real estate services; **M** Professional, scientific and technical services; **N** Administrative and support services; **O** Public administration and defence services; compulsory social security services; **P** Education services; **Q** Human health and social work services; **R** Arts, entertainment and recreation services; **S** Other services.

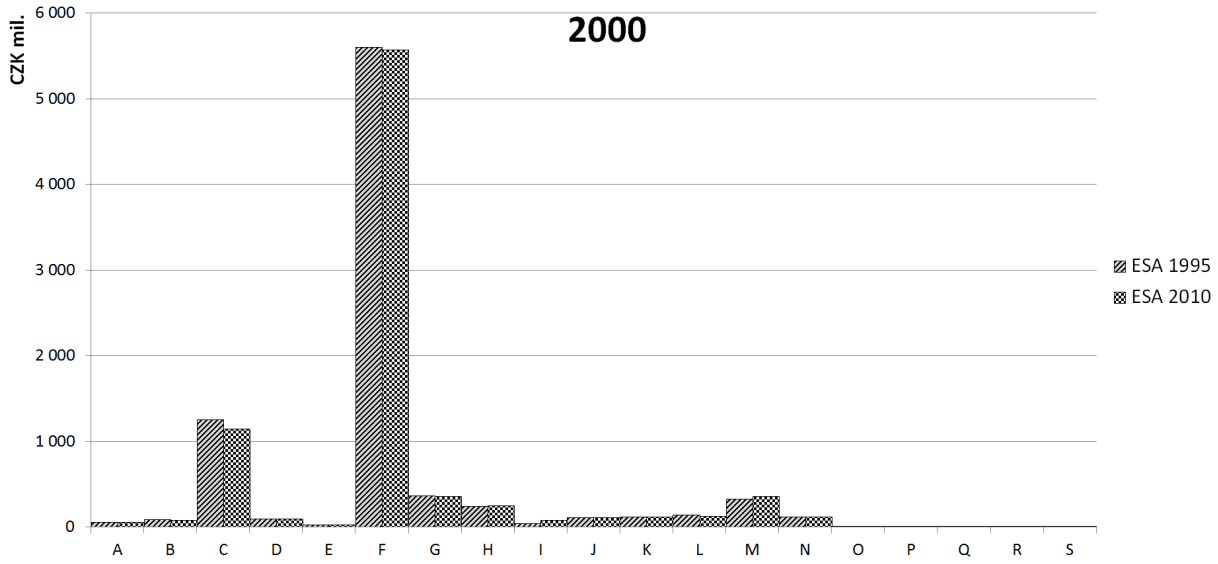


Figure 2 Comparison of results of input-output analysis based on SIOT according to ESA 1995 and ESA 2010 for 2000

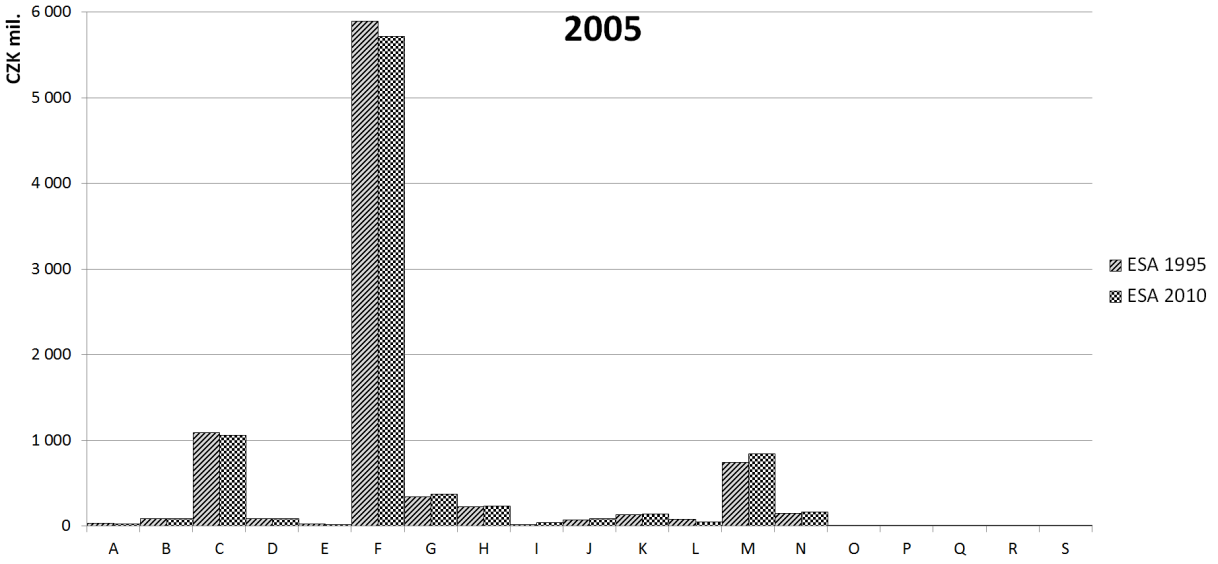


Figure 3 Comparison of results of input-output analysis based on SIOT according to ESA 1995 and ESA 2010 for 2005

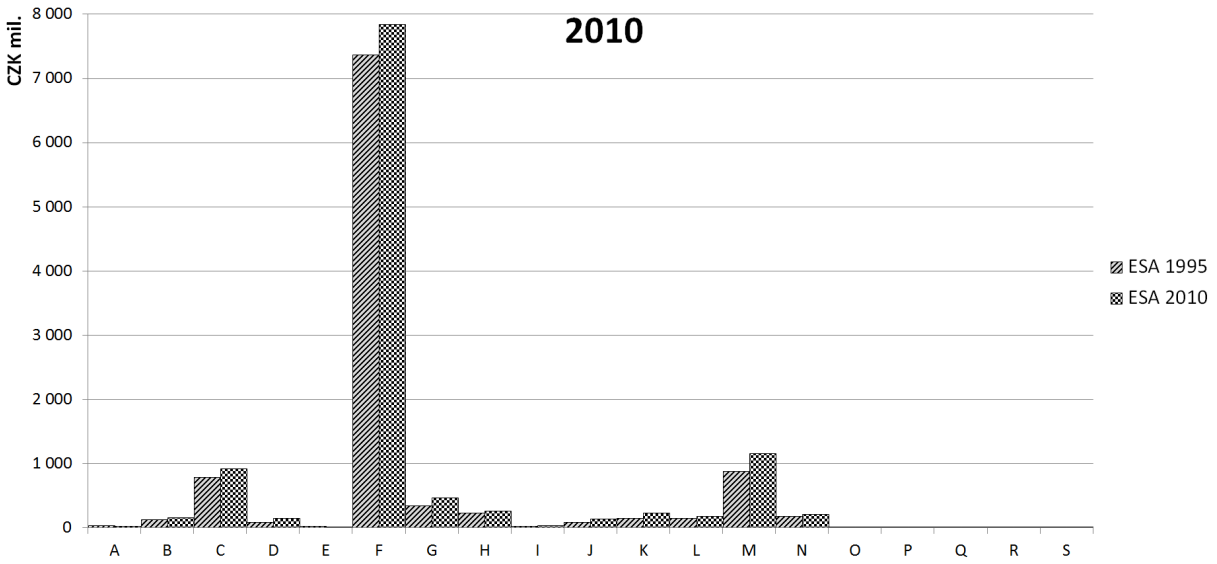


Figure 4 Comparison of results of input-output analysis based on SIOT according to ESA 1995 and ESA 2010 for 2010

scientific and technical services (*M*, 31.7%). Minor differences are evident in the case of induced production of the manufacturing or retail and wholesale products (*G*, 38.1%). Minor differences can also be seen in the production of some other products, but their level is insignificant for both statistical standards as a methodical basis. The total relative change between results for output based on both standards in this year is +13.0%.

4 Conclusion

The implementation of new statistical standards usually means complication for the users. In other words, the effort for a complete coverage of statistical agencies cause breaks in time series. Even if there is a tradition of national accountants reconstruction of time series and transforming original data into a new methodology, some evens did not existed in the past, see Sixta and Vltavská [7]. There has to be always a pressure of academic sphere on official statistical agencies otherwise the statistics will be rather used for administrative and regulatory purposes than for the description of the development of the society and research.

The results of the input-output analysis showed that the figure do not dramatically differ when using ESA 1995 or ESA 2010 Standard. It reflects the stability of the relation between output and intermediate consumption expressed in input coefficients. Since the tests were performed on IOA for the most popular case of investment shocks, there could be different results in case of change in final household consumption expenditures. In our case, the updated national accounts standards resulted to change of total output +0.3% in 1995, −1.0% in 2000, −0.4% in 2005 and +13.0% in 2010.

The consistency between researchers expectations and possibilities of modern statics is not always optimal. Current economy is changing very fast and in some cases statistics is being updated irregularly. This represent a compromise between accuracy and comparability of times series of statistical data. In the area of input-output tables, updated national accounts standards brought significant changes whose impacts were not very well estimated. The next update of national accounts standard is going to be discussed soon and some imperfections can be corrected.

Acknowledgements

Supported by the grant No. VŠE IGS F4/28/2017 and Institutional Support for Long Period and Conceptual Development of Research and Science at Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Hronová, S., Fischer, J., Hindls, R., Sixta, J.: *Národní účetnictví. Nástroj popisu globální ekonomiky*. C. H. Beck, Prague, 2009.

[2] Kramulová, J., Musil, P., Zeman, J., Michlová, R. : Regional Price Levels in the Czech Republic – Past and Current Perspectives. *Statistika* [online] **3** (2016), p. 22–34. ISSN 0322-

788X. URL: https://www.czso.cz/documents/10180/32912822/32019716q3022.pdf/c379c20d-62ee-446d-9867-05d81129a490?version=1.0.

[3] Miller, R. E. and Blair, P. D.: *Input-Output Analysis: Foundations and Extensions*. Cambridge University Press, New York, 2009. ISBN-13 978-0-511-65103-8.

[4] Musil, P., Cihlář, M.: Impact of the Implementation of ESA 2010 on Volume Measurement. *Statistika* [online] **2** (2016), p. 5–14. ISSN 0322-788X. URL: https://www.czso.cz/documents/10180/32912824/32019716q2005.pdf/59dba687-40d0-41ae-bbdd-5ba77e0bad67?version=1.1.

[5] Sixta, J., Šimková, M., Musil, P.: Implementation of ESA 2010/SNA 2008 into Czech Input-Output Tables. In: *International Input-Output Conference (IIOA)* [online]. Lisboa, 14.07.2014 – 18.07.2014. Lisboa: University of Lisboa, 2014, p. 1–5. URL: https://www.iioa.org/conferences/22nd/papers/files/1432_20140315100_Paper_Sixta_Musil_Simkova.pdf.

[6] Sixta, J.: Development of Input-Output Tables in the Czech Republic. *Statistika* [online] **2** (2013), p. 4–14. ISSN 0322-788X. URL: http://www.czso.cz/csu/2013edicniplan.nsf/engt/9600346C97/$File/180213q2004.pdf.

[7] Sixta, J., Vltavská, K.: Změny v měření ekonomiky a dopady do odhadu produktivity. *Politická ekonomie* [online] **3** (2016), p. 351–368. Text in Czech. ISSN 0032-3233. URL: https://www.vse.cz/polek/download.php?jnl=polek&pdf=1071.pdf.

[8] Šafr, K.: Pilot Application of the Dynamic Input-Output Model. Case Study of the Czech Republic 2005–2013. *Statistika* [online] **2** (2016), p. 15–31. ISSN 0322-788X. URL: https://www.czso.cz/documents/10180/32912824/32019716q2015.pdf/c554c8bd-f80b-4cfa-985c-2516c87b428d?version=1.1.

Spatial analysis of the population aging phenomena in European Union

Katarzyna Zeug-Żebro[1], Monika Miśkiewicz-Nawrocka[2]

Abstract. The population aging phenomena is defined as changes in the age structure of population where the percentage of older population compared to the total population number is increasing. Aging process is now one of the most important economic problems in the world. The analysis of this process can be used to determine the right direction of change, such as the conduct of a family-friendly policy.
Spatial methods are used increasingly frequently in the analysis of economic processes. One of the reason is the fact that spatial autocorrelation local and global measures, informing about the type and strength of spatial dependency, allow fuller use of the measure to determine the relationship between reference entities and to define spatial structures.
The main objective of this paper is to study the spatial dependences of old-age dependency ratio in European Union using measures of global and local spatial autocorrelation. The data used in analysis come from the Eurostat. All calculations and maps will be made in the R CRAN.

Keywords: the process of aging, aging measures, spatial autocorrelation, global and local statistics.

JEL Classification: C49
AMS Classification: 62H11

1 Introduction

Population aging is a global phenomenon today, affecting all regions and almost all countries in the world. However, this process goes on very uneven. Deepest in Europe, where the lowest birth rate has been observed since almost the beginning of the last century. As a consequence, Europe is demographically the oldest continent. The low birth rate and the relatively fast aging of Europe are due to a number of economic and socio-cultural factors characteristic of highly developed societies. First of all, they include life-longening factors, such as modern health care systems and the care of older people, and secondly those that affect the reduction of procreation, such as long periods of education or high levels of professional activity.

The aging process is a multifaceted phenomenon affected by, inter alia, the demographic and spatial nature of individuals. The application of spatial analysis to investigate this process will allow to establish existing relationships between the studied regions with respect to this phenomenon. It will allow not only general but also individual characteristics of similarities and differences of regions.

The main objective of this paper is to study the spatial dependences of old-age dependency ratio in European Union using measures of global and local spatial autocorrelation. All calculations and maps were made in the statistical program R CRAN and Microsoft Excel. The data was obtained from the Eurostat [www 1], 2001-2016.

2 Analysis of aging process in European Union

The old-age dependency ratio OADR is one measure that, alongside the demographic aging index and median age, is very often used to assess the extent of the aging process. This factor is defined as the ratio of the number of people 65 years or older to the number of people ages 15 through 64. Since 1950 factors OADR have grown steadily in all regions of the world, and according to forecasts published in the paper UN [11], this increase will continue in the future. Faster growth rates can be expected in high and medium developed regions. It is worth stressing that in the highly developed countries, this growth will not be strongest, as these are the most advanced countries in terms of population aging and thus already characterized by quite high values of this coefficient. It is worth stressing that in the highly developed countries, this growth will not be strongest, as these are the most

[1] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, katarzyna.zeug-zebro@ue.katowice.pl.
[2] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, monika.miskiewicz@ue.katowice.pl.

advanced countries in terms of population aging and thus already characterized by quite high values of this coefficient.

Since Europe is the oldest (in a demographic sense) region in the world, the results of the estimation of old-age dependency ratio for individual regions (NUTS1) of the European Union should be presented. The obtained results were the basis for the classification of these regions, due to homogeneous groups, from the point of view of the achieved degree of the studied phenomenon (ie the level of advancement of the aging process). The total volatility range of measures was divided into four class to which the regions were assigned according to the following rules [14]:

- class I (high degree of aging): $\overline{OADR} + S_{OADR} \leq OADR_i$,
- class II (average degree of aging): $\overline{OADR} \leq OADR_i < \overline{OADR} + S_{OADR}$,
- class III (low level of aging): $\overline{OADR} - S_{OADR} \leq OADR_i < \overline{OADR}$,
- class IV (very low level of aging): $OADR_i < \overline{OADR} - S_{OADR}$,

where

$$\overline{OADR} = \frac{1}{n}\sum_{i=1}^{n} OADR_i \ , \ S_{OADR} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(OADR_i - \overline{OADR}\right)^2} \ . \tag{1}$$

The results of the spatial distribution of the designated classes in 2001, 2009 and 2016 are shown in the following maps (Figure 1):

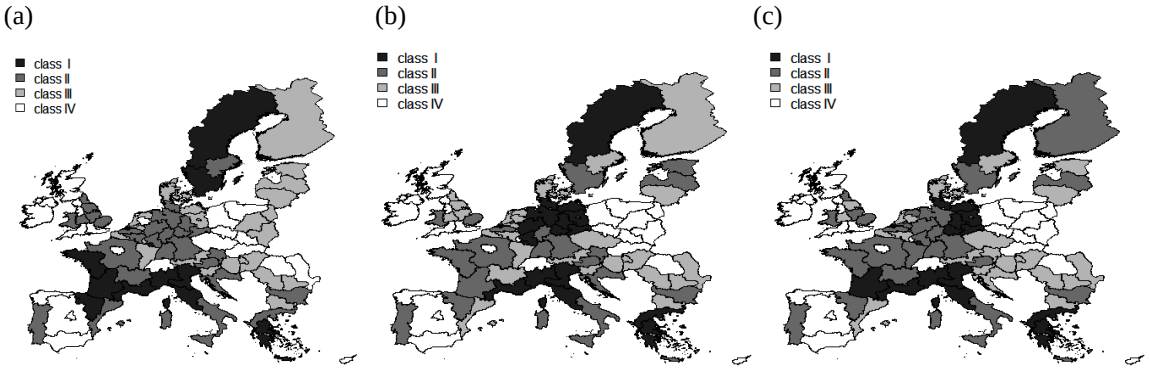


Figure 1 Spatial distribution of regions of the European Union according to level of the old-age dependency ratio in (a) 2001, (b) 2009 and (c) 2016

Analyzing these maps it can be noted that in the years 2001, 2009 and 2016, half of the regions of European Union belonged to the class I or II which indicate the high and medium level of aging. During the period under review, 59 spatial units didn't change their place in the classification, 17 regions changed the class in which the studied phenomenon was higher and in 22 units changed class to one which was characterized by a lower level of aging. The regions with the lowest values of OADR were scattered throughout the European Union, but most of them were located in the eastern part (Poland, Slovakia, Cyprus and part of Romania) and western (Spain, Ireland, northern and southern regions of UK). On the other hand, the highest values of this measure concerned primarily areas, initially Sweden, western and southern France, Greece and northern regions of Italy, later also Germany.

3 Spatial autocorrelation

The spatial autocorrelation occurs when a certain phenomenon in a single spatial unit alters the probability of occurrence of this phenomenon in the neighboring units [3,12]. In general, the positive spatial autocorrelation occurs when we observe the accumulation, in terms of the location, high or low values of observed variables. In the case of negative autocorrelation, high values adjacent to low, and low to high, creating a kind of checkerboard [10]. The lack of spatial autocorrelation means the spatial randomness, i.e. the high and low values of observed variables are distributed independently.

3.1 Global Moran's *I*

The Moran statistics [8] is one of the most widely used measures in the study of spatial autocorrelation. The Global Moran's I is defined as follows:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{n}{S_0} \cdot \frac{z^{\mathrm{T}} W z}{z^{\mathrm{T}} z} \tag{2}$$

where: x_i, x_j are the values of variables in spatial unit *i* and *j*, $\bar{x}$ is the mean of variable for all units, *n* is the total number of spatial units that are included in the study, S_0 is the sum of all elements of a spatial weight matrix, *z* is a column vector of elements $z_i = x_i - \bar{x}$, *W* is the spatial weight matrix degree *n*, defining the structure of the neighborhood, w_{ij} is an element of weights matrix *W* [7]. This statistic takes values ranging from $[-1,1]$: positive, when tested objects are similar, negative, when there is no similarity between them and approximately equal to 0 for a random distribution of objects.

Cliff and Ord [4] have shown that the distribution of Moran statistics is asymptotically normal. Thus, the statistical significance of spatial autocorrelation can be verified using normalized statistics : $I_S \sim N(0,1)$:

$$I^S = \frac{I - \mathrm{E}(I)}{\sqrt{\mathrm{Var}(I)}} \tag{3}$$

where: $\mathrm{E}(I)$ is the expected value of Moran's and $\mathrm{Var}(I)$ is its variance:

$$\mathrm{E}(I) = -\frac{1}{n-1},\ \mathrm{Var}(I) = \frac{n^2 S_1 - nS_2 + 3S_0^2}{(n^2-1)S_0^2} - \frac{1}{(n-1)^2} \tag{4}$$

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\ ,\ S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(w_{ij} + w_{ji}\right)^2,\ S_2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2 \tag{5}$$

If the Moran statistic has a value $I \approx -(n-1)^{-1}$, $I^S \approx 0$ it indicates a random spatial pattern. However, when $I > -(n-1)^{-1}$, $I^S > 0$ the spatial autocorrelations is positive, and if $I < -(n-1)^{-1}$, $I^S < 0$, the spatial autocorrelations is negative.

3.2 Global Geary's *C*

Another global measure of spatial autocorrelation is Global Geary's *C*[5]. This statistic, is given by

$$C = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - x_j\right)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{n}{(n-1)}\left[\frac{n}{S_0} \cdot \frac{z^{\mathrm{T}}\mathrm{diag}(w_{i.})z}{z^{\mathrm{T}} z} - I\right] \tag{6}$$

where all elements of the formula are defined as in statistic *I*. The above formula shows that the Geary measure can be expressed by the Moran statistic [6]. Although Moran and Geary measures give similar results, the Moran statistic is more effective. This is due to greater sensitivity of the variance of the Geary statistic to the distribution of sample. Values of this statistic can be impaired when the matrix of weights is asymmetrical. In order to verify the hypothesis of no spatial correlation, the Geary statistic can be standardized:

$$C^S = \frac{C - \mathrm{E}(C)}{\sqrt{\mathrm{Var}(C)}} \sim N(0,1) \tag{7}$$

where: $\mathrm{E}(C)$ is the expected value of Geary's and $\mathrm{Var}(C)$ is its variance:

$$\mathrm{E}(C) = 1,\ \mathrm{Var}(C) = \frac{(n-1)(2S_1 + S_2) - 4S_0^2}{2(n+1)S_0^2} \tag{8}$$

The value of Global Geary's *C* is always positive and takes values ranging from $[0,2]$. In the case, of: $1 < C < 2$, $C^S > 0$, the spatial autocorrelation is negative; when $0 < C < 1$, $C^S < 0$, the spatial autocorrelation is positive; finally, when $C \approx 1$, $C^S \approx 0$, there is no spatial autocorrelation.

3.3 Local Moran's I_i

The Local Moran determines clusters of spatial units and studies whether the unit is surrounded by neighboring units with similar or different values of the variable studied in relation to the random distribution of these values in the studied space [7].

In the case of non-standardized values of the variable and row-standardized spatial weight matrix [2] ($\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} = n$), the local Moran is given by:

$$I_i = \left[(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij} (x_j - \bar{x}) \right] \Big/ \left[\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} \right] \tag{9}$$

where all elements of the formula are defined as in the Global Moran's *I*. The standardized Local Moran's I_i^S is used to test the statistical significance of local spatial autocorrelation [1]:

$$I_i^S = \frac{I_i - \mathrm{E}(I_i)}{\sqrt{\mathrm{Var}(I_i)}} \sim N(0,1) \tag{10}$$

where: $\mathrm{E}(I_i)$ is the expected value of the Local Moran and $\mathrm{Var}(I_i)$ is its variance

$$\mathrm{E}(I_i) = -\frac{\sum_{j=1}^{n} w_{ij}}{n-1} \text{ i } \mathrm{Var}(I_i) = \frac{(n-k)\sum_{i\neq j} w_{ij}^2}{n-1} + \frac{2(2k-n)\sum_{l\neq i}\sum_{hi\neq i} w_{il} w_{ih}}{(n-1)(n-2)} - \left(\frac{-\sum_{i\neq j} w_{ij}}{n-1} \right)^2 \tag{11}$$

where $k = \left(\frac{1}{n}\sum_i (x_i - \bar{x})^4 \right) \Big/ \left(\frac{1}{n}\sum_i (x_i - \bar{x})^2 \right)^2$.

When I_i^S is negative, the spatial autocorrelation is negative too, i.e. when the object is surrounded by spatial units with significantly different values of the studied variable. The spatial autocorrelation is positive when $I_i^S > 0$, the object is surrounded by similar neighboring units.

According to Anselin [1] a Local Geary statistics for an observation *i* may be defined as

$$C_i = \sum_{j\neq i}^{n} w_{ij} (z_i - z_j)^2 \tag{12}$$

where $z_i = x_i - \bar{x}$, $z_j = x_j - \bar{x}$ and w_{ij} are the elements of the row-standardized binary symmetric spatial weight matrix **W**. The test statistic for C_i^S is

$$C_i^S = \frac{C_i - \mathrm{E}(C_i)}{\sqrt{\mathrm{Var}(C_i)}} \sim N(0,1) \tag{13}$$

where: $\mathrm{E}(I_i)$ is the expected value of the Local Moran and $\mathrm{Var}(I_i)$ is its variance

$$\mathrm{E}(C_i) = \frac{n\sum_{j=1}^{n} w_{ij} \cdot \sum_{j=1}^{n} (z_i - z_j)^2}{(n-1)^2} \text{ i } \mathrm{Var}(C_i) = \frac{\left[(n-1)\sum_{i=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij} \right)^2 \right] \cdot \left[(n-1)\sum_{j=1}^{n} (z_i - z_j)^4 - \left[\sum_{j=1}^{n} (z_i - z_j)^2 \right]^2 \right]}{(n-1)^2 (n-2)} \tag{14}$$

the significant testing on local spatial association can be conducted based on the calculated test statistics above. The C_i statistic is interpreted in the same way as the Local Moran.

4 The empirical analysis

The object of the study were regions (NUTS 1) of the European Union in years 2001, 2009 and 2016. In the first stage of the research estimated the spatial autocorrelation of EU's society aging process (based on the old-age dependency ratio calculated for each region). Calculated values of global Moran and global Geary are presented in Tables 1.

Year	Global Morana statistic				Global Geary statistic			
	I	E (*I*)	Var (*I*)	*p-value*	*C*	E (*C*)	Var (*C*)	*p-value*

2001	0.35696	-0.01136	0.00698	$5.22 \cdot 10^{-6}$	0.61507	1.00000	0.00764	$5.32 \cdot 10^{-6}$
2009	0.36488	-0.01136	0.00699	$3.38 \cdot 10^{-6}$	0.60013	1.00000	0.00759	$2.23 \cdot 10^{-6}$
2016	0.31891	-0.01136	0.00695	$3.72 \cdot 10^{-5}$	0.63071	1.00000	0.00792	$1.67 \cdot 10^{-5}$

Table 1 Global Moran and Geary in the years 2001, 2009 and 2016

Analyzing the data contained in Table 1 can be concluded that both global statistics are positive and statistically significant. This means the similarity of spatial units (NUTS 1) due to the level of advancement of population aging process. Graphical presentation of Moran statistics in years 2001, 2009 and 2016 is presented in Figure 2.

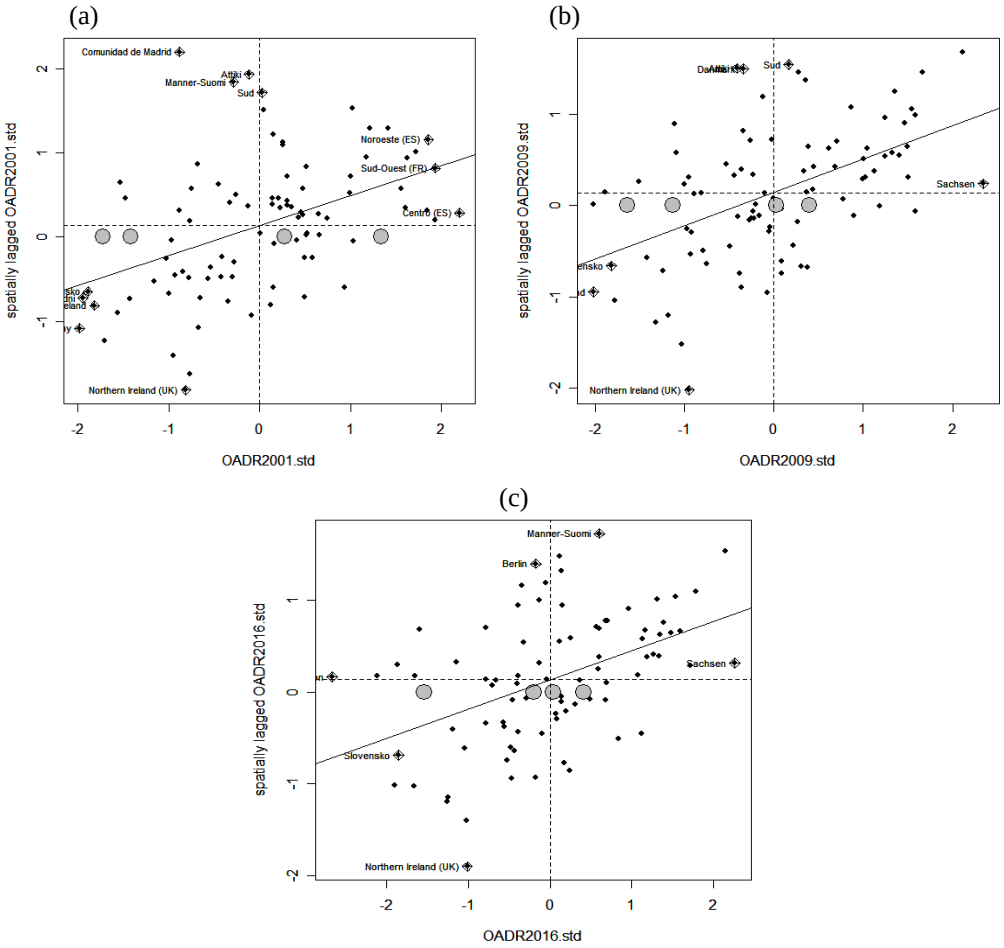


Figure 2 Scatter plot for the Moran global statistic in years: (a) 2001, (b) 2009, (c) 2016

These graphs show regions with outlying OADR values. Regions: Centro-ES4, Éire/Ireland-IE0, Region Pólnocno-Zachodni-PL4, Region Pólnocny-PL6 (2001), Slovensko-SK0, Northern Ireland-UKN (2001, 2009, 2016), Sachsen-DED (2009, 2016) are located below the regression line. The OADR values at these locations outweigh the values in neighborhood regions more than it would result from the overall spatial pattern. Conversely, the neighbors of the regions above the regression line (Noroeste-ES1, Comunidad de Madrid-ES3, Sud-Ouest-FR6 (2001), Danmark-DK0 (2009), Attiki-EL3, Sud-ITF (2001, 2009), Manner-Suomi-FI1 (2001, 2016), Berlin-DE3, London (2016)) have higher, than the mean, values of the old-age dependency ratio.

The next stage of research was to estimate the local Moran in order to identify the spatial structure. The results of spatial distribution for the regions of European Union in the years 2001, 2009, 2016 are shown in Figure 3. Based on these maps (Figure 3) it can be seen that only some of the local Moran are statistically significant. In the studied period, significant and positive values of local Moran obtained for the factor OADR only for 16 regions in 2001, 19 in 2009 and 17 in 2016. This means that in chosen years, these regions have been surrounded by units with similar values of the old-age dependency ratio which expressing the level of an aging population. Therefore, the above-mentioned regions of European Union have been clusters. For other units local Moran's I_i was statistically insignificant

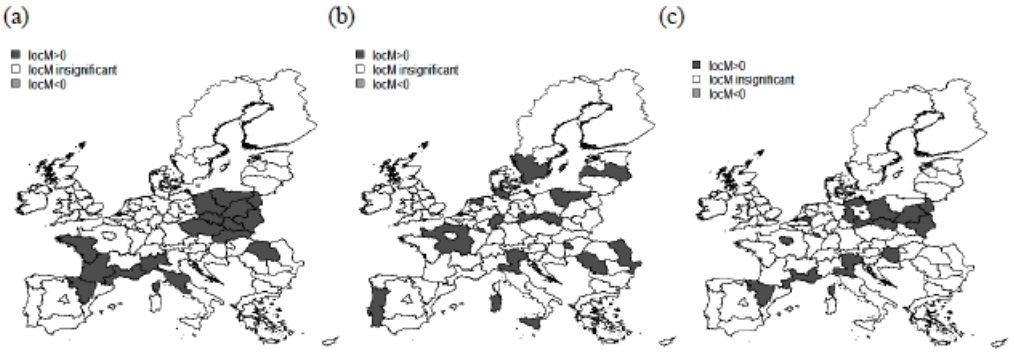


Figure 3 The classification of regions of the European Union due to the value of the local Moran for the old-age dependency ratio in years: (a) 2001, (b) 2009, (c) 2016.

5 Conclusions

Aging of population is a global and irreversible process, differentiated regionally [13]. It can be assumed that the higher the degree of socio-economic development of a country, the higher the aging of the population. This is confirmed by the classification of regions of the European Union based on the old-age dependency ratio. In the years 2001, 2009 and 2016 Sweden, the western and southern parts of France, Greece, Germany and the northern regions of Italy were areas which characterized the most advanced ageing process. On the other hand, the areas of Poland, Slovakia, Cyprus, parts of Romania and Spain, Ireland and the northern and southern regions of Great Britain were demographically younger. The spatial dependence analyzes of the European Union's aging population show that there is a positive spatial autocorrelation, ie the formation of clusters with similar OADR factor. These dependencies are also evident when considering local statistics. This analysis can therefore serve to identify areas with similar or significantly different values of the features being investigated, which allow to monitor the phenomenon and its control [9].

References

[1] Anselin, L.: Local Indicators of Spatial Association-LISA. *Geographical Analysis*, **27** (1995), 93-115.
[2] Arbia, G.: *Spatial Econometrics: Statistical Foundations and Applications to Regional Growth Convergence.* Springer, New York, 2006.
[3] Bivand, R.: Autokorelacja przestrzenna a metody analizy statystycznej w geografii. In: *Analiza regresji geografii* (Chojnicki, Z., ed), PWN, Poznań, 1980, 23-38.
[4] Cliff, A.D. and Ord, J.K.: *Spatial Autocorrelation*. Pion, London, 1973.
[5] Geary, R.: The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician* **5** (1954), 115–145.
[6] Griffith, D.A.: *Spatial Autocorrelations and Spatial Filtering*. Springer, Berlin-Heidelberg, 2003.
[7] Kopczewska, K.: *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*. Cedewu.pl, Warszawa, 2006.
[8] Moran, P. A. P.: Notes on Continuous Stochastic Phenomena. *Biometrika* **37** (1), 1950, 17–23.
[9] Pośpiech, E. and Mastalerz-Kodzis, A.: A Spatial and Temporal analysis of Labour Market Characteristics. *Folia Oeconomica Stetinensia* **16** (2), 2016, 60-71.
[10] Suchecki, B. (ed.): *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych.* Wydawnictwo C.H. Beck, Warszawa, 2010.
[11] United Nations, Department of Economic and Social Affairs, Population Division: *World Population Prospects: The 2015 Revision, Key Findings and Advance Tables.* Working Paper No. ESA/P/WP.241. 2015. Retrieved April 26, 2017, from https://esa.un.org/unpd/wpp/publications/ files/key_findings_wpp_2015.pdf.
[12] Warzecha, K.: The use of quantitative methods in research on selected behavioral addictions of young people. *Studia Ekonomiczne* **247** (2015), 121-139.
[13] Wolny-Dominiak A. and Zeug-Żebro K., Spatial statistics in the analysis of county budget incomes in Poland with the R CRAN, In: *Proceedings of 30th International Conference Mathematical Methods in Economics.* (Ramik J., Stavárek, D., eds.), Karviná: Silesian University, School of Business Administration, 2012, 992-998.
[14] Zeug-Żebro K. and Miśkiewicz-Nawrocka M.: Ocena stopnia zagrożenia bezrobociem województw Polski w latach 2005-2012, *Zeszyty Naukowe, "Cracow Review of Economics and Managment"* **4** (940), 2015, 145-161.
[15] Eurostat. Retrieved April 26, 2017, from http://epp.eurostat.ec.europa.eu/.

Oscillators and their usefulness in foreign exchange trading

Tomas Zuscak [1]

Abstract. Traders use different strategies in deciding whether to buy or sell foreign currencies. They can use various instruments of the technical analysis of exchange rates. Oscillators are one of them, e.g. an absolute or a relative momentum, a moving average convergence divergence (MACD), a relative strength index (RSI), a stochastic oscillator, a commodity channel index (CCI), Williams´ percent range and many others. This paper deals with usefulness of some favorite oscillators in foreign exchange trading. The article presents both theoretical basis and selected trading strategies based on these oscillators. The aim of this paper is to test these strategies. Using EUR/USD historical data with daily time frame, we look for signals for opening a long or a short position. Consequently, we evaluate the proportion of successful trades. In conclusion, we discuss results of these tests.

Keywords: technical analysis, foreign exchange rates, oscillators, evaluation of trading strategies.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Trading in foreign currency is very popular these days. With the boom of electronic trading systems, small traders are also involved in this activity. Traders use different strategies in deciding whether to buy or sell foreign currencies. They especially look for signals for opening a long or a short position. They can use various instruments of the technical analysis of the exchange rates for this purpose. Oscillators are one of them. This paper deals with effectivity of some favorite oscillators as the sources of these signals.

2 Oscillators

We introduce six favorite oscillators in foreign exchange trading that we use in tested strategies, i.e. an absolute momentum, a moving average convergence divergence, a relative strength index, a stochastic oscillator, a commodity channel index and Williams´ percent range.

We denote P_t the exchange rate of a currency pair in the time period t ($t \in \mathbf{Z}$). We denote C_t (resp. O_t or H_t or L_t) the closing (resp. the opening or the highest or the lowest) exchange rate of the currency pair in the time period t ($t \in \mathbf{Z}$). We use following finite time series to define the oscillators: $\{P_t, t = 1, \dots, T\}$, $\{C_t, t = 1, \dots, T\}$, $\{O_t, t = 1, \dots, T\}$, $\{H_t, t = 1, \dots, T\}$, $\{L_t, t = 1, \dots, T\}$.

2.1 Definitions

In this section we briefly present theoretical basis of specified oscillators.

The absolute momentum (AM)

The absolute momentum in the time period t is defined by the formula:

$$AM_t(n) = P_t - P_{t-n} \tag{1}$$

where n is a parameter and $t = n + 1, \dots, T$. The absolute momentum oscillates around zero. We can also apply the formula (1) on the closing, the opening, the highest or the lowest rates (i.e. Close, Open, High and Low).

The moving average convergence divergence (MACD)

The moving average convergence divergence in the time period t is defined by the formulas:

$$MACD_t(n_1, n_2) = EMA_t(P_t, n_1) - EMA_t(P_t, n_2) \tag{2}$$

[1] University of Hradec Kralove, Department of Mathematics, Rokitanskeho 62, Hradec Kralove, Czech Republic, e-mail: tomas.zuscak@uhk.cz.

$$SIGNAL_t(n_3) = EMA_t(MACD_t(n_1, n_2), n_3) \tag{3}$$

where n_1, n_2, n_3 are parameters ($n_1 < n_2$) and $t = n_2 + n_3 + 1, \ldots, T$. The symbol $EMA_t(X_t, n)$ denotes an exponential moving average of X_t with a length of n periods. $MACD_t(n_1, n_2)$ is the difference between faster and slower exponential moving averages of P_t and it is complemented by the signal curve $SIGNAL_t(n_3)$ (i.e. an exponential moving average of the MACD). Both the MACD and the signal curve oscillate around zero. We can also apply the formula (2) on Close, Open, High and Low. We denote $MACD_t(n_1, n_2, n_3)$ both the MACD and the signal curve.

The relative strength index (RSI)

The relative strength index in the time period t is defined by the formula:

$$RSI_t(n) = 100 - \frac{100}{1 + RS_t(n)} \tag{4}$$

where n is a parameter and $t = n + 1, \ldots, T$. The symbol $RS_t(n)$ denotes the relative strength of P_t in the time period t. It is defined as a ratio of an average positive price change to an average negative price change during a specified time period with a length of $n + 1$ (i.e. it is calculated by using $P_{t-n}, \ldots, P_t$). The relative strength index oscillates around 50 between 0 and 100. We can also apply the formula (4) on Close, Open, High and Low.

The stochastic oscillator (Stoch)

The stochastic oscillator in the time period t is defined by the formulas:

$$\%K_t(n) = \frac{C_t - L_n}{H_n - L_n} \cdot 100 \tag{5}$$

$$\%D_t(m) = SMA_t(\%K_t(n), m) \tag{6}$$

where n, m are parameters and $t = n + m - 1, \ldots, T$. The symbol L_n denotes the minimum of the lowest rates during a specified time period with a length of n, i.e. $L_n = \min\{L_i; i = t - n + 1; t\}$. The symbol H_n denotes the maximum of the highest rates during the same time period, i.e. $H_n = \max\{H_i; i = t - n + 1; t\}$. $\%D_t(m)$ is defined as a simple moving average of $\%K_t(n)$ with a length of m periods. Both %K and %D oscillate around 50 between 0 and 100. We denote $Stoch_t(n, m)$ both %K and %D.

The commodity channel index (CCI)

The commodity channel index (CCI) in the time period t is defined by the formula:

$$CCI_t(n) = \frac{TP_t - SMA_t(TP_t, n)}{0{,}015 \cdot MD_t(TP_t, n)} \tag{7}$$

where n is a parameter and $t = n, \ldots, T$. The symbol TP_t denotes the typical exchange rate in the time period t and it is defined as an arithmetic mean of C_t, H_t and L_t. $SMA_t(TP_t, n)$ is a simple moving average of TP_t with a length of n periods and $MD_t(TP_t, n)$ is an absolute mean deviation over the same period. The commodity channel index oscillates around zero.

Williams´ percent range (Williams %R)

Williams´ percent range in the time period t is defined by the formula:

$$\%R_t(n) = -\frac{H_n - C_t}{H_n - L_n} \cdot 100 \tag{8}$$

where n is a parameter and $t = n, \ldots, T$. The symbols L_n and H_n denote the same quantities as in the case of the stochastic oscillator. Williams´ percent range oscillates around -50 between -100 and 0. It is obvious that the following formula applies: $\%K_t(n) - \%R_t(n) = 100$.

2.2 Signals

We test four usual types of signals based on the above-mentioned oscillators to open either a long or a short position. We compare them with signals based on results of random experiments. This comparison gives us in-

teresting results. In the following text we work with a daily time period. However, our considerations can easily be applied for another time period.

The oscillator rises above a certain value

If the oscillator rises above a certain value X it is understood as a signal to open a long position. Yet we wait one more day to see whether the oscillator continues growing in order to avoid a possible false signal. If the growth is confirmed, we open the position the following day (i.e. we buy currency for tomorrow´s opening rate). We denote this signal *Rise*(X). We use these values for the individual oscillators: AM (X = 0), MACD (X = 0), RSI (X = 50), Stoch %D (X = 20), CCI (X = 0, 100, -100), Williams %R (X = -80).

In some cases, it is sufficient for us if the oscillator rises above a certain value X and we no longer wait for further confirmation of growth. Thus we open the position the next day after the day when the threshold is exceeded. We denote this signal *Rise2*(X). We use this approach for RSI (X = 30).

The oscillator falls below a certain value

If the oscillator falls below a certain value X it is understood as a signal to open a short position. If the decline is confirmed, we open the position the following day (i.e. we sell currency for tomorrow´s opening rate). We denote this signal *Fall*(X). We use these values for the individual oscillators: AM (X = 0), MACD (X = 0), RSI (X = 50), Stoch %D (X = 80), CCI (X = 0, 100, -100), Williams %R (X = -20).

If we do not require further confirmation of the decline, we denote the signal *Fall2*(X). We use this approach for RSI (X = 70).

The oscillator rises above its signal curve

Only the MACD and the stochastic oscillator give this type of a signal. If the main curve of the oscillator (MACD or %K) is under its signal curve (Signal or %D), and then it rises above its signal curve, it is understood as a signal to open a long position. We wait one more day to see whether a difference between the main curve and the signal curve continues growing. If the growth is confirmed, we open the position the following day. We denote this signal *CutL*.

The oscillator falls below its signal curve

This type of a signal is the opposite of the above-mentioned one. It is understood as a signal to open a short position. We denote this signal *CutS*.

Signals based on the result of random experiments

We test the following six types of signals. We throw a dice and if we drop 5 or 6 (resp. 1 or 2) we open a long position (resp. a short position) the following day. We denote this signal *DiceL* (resp. *DiceS*). We throw a coin and if we drop heads (or tails) we open a long position (or a short position) the following day. We denote this signal *CoinL* (resp. *CoinS*). We generate a random integer between 1 and n and if it equals n (or 1), we open a long position (or a short position) the following day. We denote this signal *RndL*(n) (resp. *RndS*(n)).

3 Test methods and data

We test a success rate of the signals using the EUR/USD historical data with daily time frame that are freely available at **https://www.investing.com/currencies/eur-usd-historical-data**. We use daily Close, Open, High and Low from April 2, 2007 to March 30, 2017 (i.e. 2643 daily records). We look for signals to open long or short positions. We evaluate success of each signal using two different methods.

Firstly, we suppose that we buy or sell 1 EUR for a price in USD and we choose the amount of profit in USD (*TakeProfit*), and the length of the period (*Duration*) during which we want to achieve the profit. We consider the signal as successful if *TakeProfit* is achieved in the course of *Duration*. If we get a signal to open a long position on the day t and we open it on the day $t + 1$ the signal can be considered as successful if the following inequality applies:

$$\max\{H_i; i = t+1,\ldots,t+Duration\} \geq O_{t+1} + TakeProfit \tag{9}$$

The signal to open a short position can be considered as successful if the following inequality applies:

$$\min\{L_i; i = t+1,\ldots,t+Duration\} \leq O_{t+1} - TakeProfit \tag{10}$$

If *TakeProfit* is achieved in the course of *Duration* we close the position immediately. If it is not be achieved we close the position at the end of *Duration*. We calculate a profit or a loss resulting from the closed position (we do not consider a spread). The profit equals to *TakeProfit* if the signal is successful. The profit or the loss equals to $C_{t+Duration} - O_{t+1}$ in the case of a long position (or $O_{t+1} - C_{t+Duration}$ in the case of a short position) if the signal is not successful. We also calculate an average income, an average profit and an average loss.

Secondly, we consider a signal to open a long (or a short) position as successful if closing rates show an increasing - bullish (or a decreasing - bearish) trend in next days. We choose the length of time period (k days) decisive to a trend assessment. If we get a signal to open the position on the day t then we calculate the slope of the regression line determined by closing rates $C_t, \ldots, C_{t+k-1}$ depending on time. The signal to open a long (or a short) position can be considered as successful if the slope is positive (or negative).

We have to set the reference value π to evaluate a success rate of each type of a signal. Let us suppose we open a long (or a short) position each day, and then we evaluate success using the first or the second method. After that we calculate success rates of long (or short) positions determined by the first and the second method and we denote them π_{L1} and π_{L2} (or π_{S1} and π_{S2}). We set π_{L1} and π_{L2} (or π_{S1} and π_{S2}) as above-mentioned reference values. Our goal is to find out whether the success rate of the signals listed in Section 2.2 is higher than the reference value. We denote p_{L1} and p_{L2} (or p_{S1} and p_{S2}) a success rate of a certain type of a signal to open a long (or a short) position determined by the first and the second method. We test a null hypothesis (H_0: $p_{L1} = \pi_{L1}$) against an alternative hypothesis (H_1: $p_{L1} > \pi_{L1}$) at the required significance level of 5 %. We calculate p-value using the formula:

$$p\text{-value} = 1 - F(NoSS) \tag{11}$$

where $F(x)$ is the distribution function of the binomial distribution $Bi(NoS, \pi_{L1})$, NoS is a number of signals and $NoSS$ is a number of successful signals (i.e. $p_{L1} = x/n$). We similarly test success rates p_{L2}, p_{S1} and p_{S2}.

4 Results

Firstly, we set *TakeProfit* = 0.008 (i.e. 80 pips), *Duration* = 10 days for the first evaluation method and k = 7 days for the second method. We calculate reference values π_{L1} = 72.10 %, π_{L2} = 50.02 %, π_{S1} = 71.11 %, π_{S2} = 49.98 %. There are results for the signals to open long positions in Table 1, and to open a short position in Table 2. We set the usual parameter values for oscillators. We denote *NoSS I* (or *NoSS II*) a number of successful signals evaluated by the first (or the second) method.

Oscillator	Signal	*NoS*	*NoSS I*	p_{L1}	*p*-value	Avg. income	Avg. profit	Avg. loss	*NoSS II*	p_{L2}	*p*-value
AM_t(14)	*Rise*(0)	73	56	76.71	15.63	0.0013	0.0080	0.0226	42	57.53	8.04
$MACD_t$(12, 26, 9)	*Rise*(0)	43	35	81.40	5.79	0.0042	0.0079	0.0145	26	60.47	6.35
$MACD_t$(12, 26, 9)	*CutL*	82	62	75.61	20.44	0.0022	0.0079	0.0179	44	53.66	22.09
RSI_t(14)	*Rise*(50)	75	56	74.67	27.02	0.0019	0.0080	0.0160	43	57.33	8.32
RSI_t(14)	*Rise2*(30)	39	25	64.10	86.61	-0.0038	0.0080	0.0249	11	28.21	99.53
$Stoch_t$(14, 3)	*Rise*(20)	75	58	77.33	12.59	0.0022	0.0079	0.0188	42	56.00	12.47
$Stoch_t$(14, 3)	*CutL*	199	145	72.86	37.87	0.0001	0.0079	0.0232	99	49.75	50.23
CCI_t(10)	*Rise*(0)	104	80	76.92	11.23	0.0023	0.0079	0.0186	62	59.62	1.97
CCI_t(10)	*Rise*(100)	87	63	72.41	43.35	0.0000	0.0078	0.0228	33	37.93	98.44
CCI_t(10)	*Rise*(-100)	136	98	72.06	47.18	-0.0007	0.0078	0.0254	70	51.47	33.58
$\%R_t$(14)	*Rise*(-80)	84	64	76.19	16.95	-0.0010	0.0078	0.0334	44	52.38	29.41
–	*DiceL*	862	621	72.04	50.23	0.0002	0.0078	0.0220	430	49.88	51.83
–	*CoinL*	1296	910	70.22	93.00	-0.0005	0.0079	0.0221	616	47.53	96.12

Table 1 Evaluation of the signals to open long positions (*TakeProfit* = 0,008, *Duration* = 10, k = 7)

We cannot prove that all tested oscillator-based signals have a success rate higher than the reference value if we want to achieve *TakeProfit* in the course of *Duration*. We can prove that only the signal *Rise*(0) based on the CCI has a success rate higher than the reference value evaluated by the second method, i.e. it could signal a bullish trend. The p-values of two signals (*Rise2*(30) based on the RSI and *Rise*(100) based on the CCI) are higher than 95 %. So, we can prove that they could signal a bearish trend.

Oscillator	Signal	*NoS*	*NoSS I*	p_{L1}	*p*-value	Avg. income	Avg. profit	Avg. loss	*NoSS II*	p_{L2}	*p*-value
AM_t (14)	*Fall*(0)	76	63	82.89	0.61	0.0028	0.0080	0.0222	41	53.59	21.01
$MACD_t$ (12, 26, 9)	*Fall*(0)	42	27	64.29	79.21	-0.0040	0.0080	0.0256	20	47.62	56.02
$MACD_t$ (12, 26, 9)	*CutS*	87	57	65.52	84.90	-0.0023	0.0080	0.0225	41	47.13	66.45
RSI_t (14)	*Fall*(50)	71	51	71.83	40.23	-0.0027	0.0080	0.0299	36	50.70	40.50
RSI_t (14)	*Fall2*(70)	21	18	85.71	3.43	0.0062	0.0079	0.0103	12	57.14	19.12
$Stoch_t$ (14, 3)	*Fall*(80)	89	66	74.16	22.85	0.0001	0.0080	0.0236	44	49.44	49.85
$Stoch_t$ (14, 3)	*CutS*	186	132	70.97	48.94	-0.0022	0.0079	0.0294	97	52.15	25.30
CCI_t (10)	*Fall*(0)	113	86	76.11	9.29	0.0021	0.0078	0.0201	58	35.19	51.33
CCI_t (10)	*Fall*(100)	131	96	73.28	26.21	0.0006	0.0079	0.0223	70	53.44	18.99
CCI_t (10)	*Fall*(-100)	96	69	71.88	39.63	-0.0009	0.0079	0.0260	60	62.50	0.51
$\%R_t$ (14)	*Fall*(-20)	85	66	77.65	7.06	0.0012	0.0080	0.0253	44	51.76	33.10
–	*DiceS*	849	617	72.67	14.83	0.0001	0.0079	0.0228	424	49.94	49.53
–	*CoinS*	1317	920	69.86	83.50	-0.0007	0.0078	0.0233	626	47.53	99.99

Table 2 Evaluation of the signals to open short positions (*TakeProfit* = 0,008, *Duration* = 10, *k* = 7)

We can only prove that the signal *Fall2*(70) based on the RSI has a success rate higher than the reference value if we want to achieve *TakeProfit* in the course of *Duration* and that the signal *Fall*(-100) based on the CCI could signal a bearish trend. Yet we cannot prove that other tested oscillator-based signals have a success rate higher than the reference value.

Secondly, we set *TakeProfit* = 0.015 (i.e. 150 pips), *Duration* = 10 days and *k* = 7 days. We calculate reference values π_{L1} = 52.43 %, π_{S1} = 49.83 %. There are results for the signals to open long and short positions in Table 3.

Oscillator	Signal	*NoS*	*NoSS I*	p_{L1}	*p*-value	Avg. income	Avg. profit	Avg. loss
AM_t (14)	*Rise*(0)	73	40	54.79	30.16	0.0026	0.0136	0.0211
$MACD_t$ (12, 26, 9)	*Rise*(0)	43	26	60.47	11.32	0.0050	0.0139	0.0151
$MACD_t$ (12, 26, 9)	*CutL*	82	46	56.10	21.93	0.0028	0.0138	0.0173
RSI_t (14)	*Rise*(50)	75	43	57.33	16.71	0.0014	0.0147	0.0186
RSI_t (14)	*Rise2*(30)	39	18	46.15	73.41	-0.0068	0.0147	0.0272
$Stoch_t$ (14, 3)	*Rise*(20)	75	42	56.00	23.17	0.0023	0.0141	0.0176
$Stoch_t$ (14, 3)	*CutL*	199	98	49.25	79.63	0.0000	0.0129	0.0217
CCI_t (10)	*Rise*(0)	104	56	53.85	34.98	0.0021	0.0138	0.0174
CCI_t (10)	*Rise*(100)	87	40	45.98	86.38	-0.0015	0.0131	0.0211
CCI_t (10)	*Rise*(-100)	136	73	53.68	35.36	-0.0005	0.0136	0.0248
$\%R_t$ (14)	*Rise*(-80)	84	47	55.95	22.53	-0.0001	0.0131	0.0297
AM_t (14)	*Fall*(0)	76	45	59.21	3.97	0.0024	0.0143	0.0205
$MACD_t$ (12, 26, 9)	*Fall*(0)	42	18	42.86	77.30	-0.0032	0.0135	0.0216
$MACD_t$ (12, 26, 9)	*CutS*	87	42	48.28	57.24	-0.0019	0.0140	0.0219
RSI_t (14)	*Fall*(50)	71	32	45.07	75.27	-0.0037	0.0135	0.0235
RSI_t (14)	*Fall2*(70)	21	12	57.14	18.74	0.0063	0.0136	0.0084
$Stoch_t$ (14, 3)	*Fall*(80)	89	48	53.93	18.95	0.0006	0.0139	0.0220
$Stoch_t$ (14, 3)	*CutS*	186	100	53.76	12.58	-0.0010	0.0138	0.0258
CCI_t (10)	*Fall*(0)	113	58	51.33	34.01	0.0003	0.0130	0.0213
CCI_t (10)	*Fall*(100)	131	65	49.62	48.44	-0.0001	0.0133	0.0213
CCI_t (10)	*Fall*(-100)	96	50	52.08	29.34	-0.0001	0.0134	0.0225
$\%R_t$ (14)	*Fall*(-20)	85	48	56.47	9.11	0.0020	0.0137	0.0204

Table 3 Evaluation of the signals to open long and short positions (*TakeProfit* = 0,015, *Duration* = 10, *k* = 7)

We can only prove that the signal *Fall* (0) based on the AM has a success rate higher than the reference value if we want to achieve *TakeProfit* in the course of *Duration*. However, the fact that at the level of significance of 5 % we cannot prove the difference between success rates and the reference values in the vast majority of cases, leads us to the next experiment. We use the signal *RndL*(30) and we generate 10 sets of results. These results are presented in Table 4.

NoS	***NoSS I***	p_{L1}	***p*-value**	**Avg. income**	**Avg. profit**	**Avg. loss**	***NoSS II***	p_{L2}	***p*-value**
86	62	72.09	45.98	0.0014	0.0078	0.0170	38	44.19	83.50
96	68	70.83	57.11	-0.0019	0.0080	0.0260	54	56.25	9.28
96	77	80.21	2.63	0.0026	0.0077	0.0245	54	56.25	9.28
79	60	75.95	18.83	0.0005	0.0079	0.0245	40	50.63	41.25
81	56	69.14	68.61	-0.0001	0.0078	0.0212	44	54.32	18.81
75	52	69.33	66.31	-0.0010	0.0080	0.0225	38	50.67	41.01
100	74	74.00	30.05	-0.0004	0.0079	0.0254	54	54.00	18.52
100	78	78.00	7.39	0.0010	0.0079	0.0252	54	54.00	18.52
82	55	67.07	81.48	-0.0035	0.0079	0.0280	46	56.10	11.28
78	56	71.79	48.11	0.0006	0.0078	0.0200	43	55.13	15.49

Table 4 Evaluation of the signals *RndL*(30) to open long positions (*TakeProfit* = 0,008, *Duration* = 10, *k* = 7)

It seems as if chance-based signals can lead us to similar success rates, average profits or losses as oscillator-based signals.

5 Conclusions and further research

We see that at the 5% level of significance we cannot prove that the success rates of strategies based on oscillator signals differ from the reference values in the vast majority of cases. Even the random selection of the moments of opening positions seems to lead to similar results. However, this observation would deserve further research. It is possible that the use of single oscillator signals does not increase the chances to open successful trades and we would be equally successful if we opened our positions randomly. But since people need some support for decision making, oscillators could help them as a placebo. Due to the relatively large average losses of unsuccessful signals, it is necessary to supplement the tested strategies with an appropriate stop loss value.

In our further research we will focus on the interrelationships between signals based on different oscillators. However, studying the success rates of multiple oscillator-based signals would require even more data records (e.g. with an hourly frame).

Finally, let us note that some of the traders are of the opinion that the moment when we open the position is not as important as the moment when we close it. The evaluation of usefulness of oscillators as sources of signals for closing positions can be an interesting direction for further research.

References

[1] Murphy, J. J.: *Technical analysis of the financial markets*. New York Institute of Finance, Paramus, 1999.
[2] Durčáková, J., Mandel, M.: *Mezinárodní finance*. Management Press, Praha, 2010.
[3] Brada, J.: *Technická analýza*. Vysoká škola ekonomická v Praze, Praha, 2000.
[4] Říha, J.: *Technická analýza cenných papírů*. Comenia Nova, Praha, 1994.
[5] Víšková, H.: *Technická analýza akcií*. HZ Systém, Praha, 1997.

Use of Malmquist Index in Evaluating Performance of Companies in Cluster

Miroslav Žižka[1]

Abstract. The article deals with the influence that a membership of a company in a cluster has on the company´s financial performance. The research was conducted on a sample of 17 companies that are members of CLUTEX, the cluster of technical textiles. For evaluating the performance, there were used number of employees, total assets and the size of capital employed as inputs. The output and the measure of financial performance was an economic value added relative to performance. The economic value added was constructed in a modular way following the methodology of the Ministry of Industry and Trade. Data was collected within the period of 2009 to 2014. Using DEA models, there was determined a rate of technical efficiency for both the assumption of a constant and the variable returns to scale for each company. The model with a focus on inputs was used. With the help of Malmquist index, a total factor productivity change and its division into a technological change, technical efficiency change and scale efficiency change was calculated for each company. According to the extent of technical and technological efficiency changes, the companies were divided into four quadrants - technological innovators, efficient - technologically backward, inefficient - innovative and inefficient – falling behind.

Keywords: Malmquist index, cluster of technical textiles, economic value added, technical efficiency change, technological change, total factor productivity change.

JEL Classification: C61, L25, L67
AMS Classification: 90B90, 90C90

1 Introduction

Clusters are currently being used as one of the instruments of economic policy to enhance the competitiveness of companies and regions. The term cluster refers to the geographic concentration of interconnected companies and institutions in a particular field. The clusters include gathering of related industries and other entities important in terms of competition. For example, they include suppliers, also the distribution channels, customers, manufacturers of complementary products and companies providing knowledge and technology. In many clusters, there are involved government and other institutions, such as universities, professional training providers, trade associations that provide specialized training, education, information, research and technical support [12]. Clusters can occur naturally or as a result of cluster initiative of a certain government institution.

The paper deals with CLUTEX, a cluster of technical textiles, which was created in 2006 as a result of a cluster initiative. The aim of this paper is to evaluate the influence of this cluster on the performance of member companies through Malmquist index and its decomposition in the period between 2009 and 2014. As a measure of performance, economic value added (EVA) was chosen, or rather the ratio index derived from it.

2 Literature review

In the first part of the research the principle of Malmquist index (hereinafter MI) is explained. The following are selected examples of the research based on MI application.

2.1 Malmquist index

In 1953 Malmquist suggested approach comparing inputs in two time periods and seeking maximum possible decrease in input which still enables the company to produce the same volume of output in the second time period. In 1982 Caves extended Malmquist input index by index of productivity. The index measures the change in productivity between t and $t + 1$ period. Another modification of the index enabled to measure the change in technical efficiency and the shift of efficient frontier for the specific production unit (DMU) [15].

MI enables to break down the efficiency change in time into two components. The first component is the change in the relative efficiency of the unit to a set of remaining units. The second component characterizes a

[1] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec 1, miroslav.zizka@tul.cz.

change in the production possibility frontier or in other words, the impact of technological change, which affects all units. When technical progress is involved, there is a shift upwards in the production possibility frontier. MI can be expressed as the geometric mean of the two components. It may be oriented to the inputs or outputs, with the assumption of constant or variable returns to scale. In this article, input oriented M_q index which can be expressed by equation (1) was used; see for example [9].

$$M_q(x^{t+1}, y^{t+1}, x^t, y^t) = E_q T_q \tag{1}$$

x_t ... inputs in t period,
y_t ... outputs in t period,
x_{t+1} ... inputs in $t + 1$ period,
y_{t+1} ... outputs in $t + 1$ period.
E_q ... change in relative efficiency of q unit with respect to other units between t and $t + 1$ period,
T_q ... change in production possibility frontier as a result of technology development between period of t and $t + 1$, or in other words technological change.

Components E_q and T_q are given by equations (2), and (3).

$$E_q = \frac{D_q^{t+1}(x^{t+1}, y^{t+1})}{D_q^t(x^t, y^t)} \tag{2}$$

$$T_q = \sqrt{\frac{D_q^t(x^{t+1}, y^{t+1}) D_q^t(x^t, y^t)}{D_q^{t+1}(x^{t+1}, y^{t+1}) D_q^{t+1}(x^t, y^t)}} \tag{3}$$

$D_q^t(x^t, y^t)$ ... production unit U_q efficiency given the existing technology in period t with inputs and outputs from period t,
$D_q^t(x^{t+1}, y^{t+1})$ ... production unit U_q efficiency given the existing technology in period t with inputs and outputs from period $t + 1$,
$D_q^{t+1}(x^t, y^t)$ ... production unit U_q efficiency given the existing technology in period $t + 1$ with inputs and outputs from period t,
$D_q^{t+1}(x^{t+1}, y^{t+1})$ ... production unit U_q efficiency given the existing technology in period $t + 1$ with inputs and outputs from period $t + 1$.

Values $D_q^t(x^t, y^t)$, $D_q^t(x^{t+1}, y^{t+1})$, $D_q^{t+1}(x^t, y^t)$, $D_q^{t+1}(x^{t+1}, y^{t+1})$ can be obtained by solving DEA models, either for the assumption of constant or increasing returns to scale.

If the value of E_q is greater than 1, it means that the company improves its technical efficiency, or catches up with other companies. In case of unit values, there was no change in technical efficiency, and if the value of E_q is less than one, then the company worsens in technical efficiency and moves away from the efficient frontier. Similarly, the values T_q can be interpreted in a way that if the index value is greater than 1, progress (innovation) in technology can be stated, if it is equal to one, no change in technology occurred in the company, and if the index is less than 1, regression in technology occurred.

Change in technical efficiency E_q can be further broken down into pure technical efficiency change PE_q and scale efficiency change SC_q. Scale efficiency SC_q change in equation (4) is the geometric mean of two measures of efficiency. The first relates to the period with t technology and the second period with $t + 1$ technology. Subscripts c, or v, denotes constant or variable returns to scale [4].

$$E_q = PE_q SC_q = \frac{D_{qv}^{t+1}(x^{t+1}, y^{t+1})}{D_{qv}^t(x^t, y^t)} \sqrt{\frac{D_{qv}^t(x^{t+1}, y^{t+1}) / D_{qc}^t(x^{t+1}, y^{t+1})}{D_{qv}^t(x^t, y^t) / D_{qc}^t(x^t, y^t)} \frac{D_{qv}^{t+1}(x^{t+1}, y^{t+1}) / D_{qc}^{t+1}(x^{t+1}, y^{t+1})}{D_{qv}^{t+1}(x^t, y^t) / D_{qc}^{t+1}(x^t, y^t)}} \tag{4}$$

2.2 Selected applications of Malmquist index

In the literature, there are many examples of the use of MI in evaluating efficiency of companies in various industries. For example, Brožová and Boháčková [2] measured efficiency change in the agricultural sector in 13 regions of the Czech Republic for the period of 2008-13 using MI. They found a decrease in efficiency in the agricultural sector within the periods of 2008/9 and 2009/10, followed by the growth of efficiency in the period

of 2010/11 and practically constant development over the next three years. This development was mainly due to technological changes during the crisis period and thereafter.

The evaluation of natural dairy cluster in China over the period of 2007-14 showed that the productivity of the dairy industry in China is not high; only 8 out of 31 clusters are efficient. Inputs were total assets, number of employees and the volume of raw milk; performance and profit were chosen as outputs. However, the overall productivity has a growing trend and is driven mainly by technological development [10].

In a study of a Spanish car market [8], the influence of product innovation on the competitive position of each model was investigated. Individual car brands were evaluated using one input (the price of the car) and seven outputs characterizing the properties of cars (ecology, consumption, performance, luggage space, safety, reliability and standard equipment). In the first phase, the position of each brand from the best-buy frontier was evaluated using DEA model with the assumption of variable returns to scale. In the second phase, there was observed the change in the brand's position towards the frontier and at the same time a shift of the frontiers marking technological advancement or regression. The research showed that a shift of efficient frontier occurred in all brands, but not equally. The biggest shift was observed in the semi-premium brands where a product innovation and lower prices were the most intense. Řepková [13] investigated the development of efficiency in the Czech banking sector between the years 2003-12. Using MI it was found that the Czech commercial banks increased their efficiency by an average of 1% per year. This change was mainly due to improvements in operational efficiency and management (by 8.5%). Conversely, the technology had a negative effect on efficiency change (decrease by 6.9%) and economies of scale had virtually no effect on efficiency change.

In a study of Ferreira et al. [7] the effect of different airports management models on their performance with regard to their location was examined. Research sample included 145 airports on three continents (Europe, Asia, and North America). The airports were divided into several categories, by continent, by the management model (individual, holding). The results showed that the biggest technological shift was observed in European airports. On the contrary, North American individually controlled airports showed the worst productivity. Overall, individually controlled airports regardless of the continent had a slightly higher productivity than the airports managed within the holding.

3 Data and methodology

The basic source of data were the financial statements (balance sheet, profit and loss statement) of the member organizations of the CLUTEX cluster for the period of 2009-14 which were obtained from MagnusWeb database [1]. Data on the number of employees were also found in this database. The selected period was chosen in the light of a cluster development and the beginning of its activities (2007), also taking into account the fact that the effects of accession to the cluster can be expected with a certain delay. Most companies have not released their financial results in the commercial register for the year 2015 yet, so the time series ends in 2014.

Research process can be divided into the following steps.

1. **Making a list of companies that form the cluster core** - the CLUTEX cluster has currently (January 2017) 33 members with heterogeneous activities - from textile production through retail activities, research up to education. Only companies with the same or a similar line of business which form the core of the cluster can be compared. These firms can be considered homogeneous production units. Based on the analysis of the subject matter according to the statistical classification NACE, these are firms operating in sectors 13200, 13900 and 14100. There are 19 such members. Financial statements of MagnusWeb database for the years 2009 to 2014 were acquired for those companies. In the case of two companies, however, obtaining reports for all monitored years failed. Therefore, the further analysis was performed on a sample of 17 companies.
2. **Defining inputs and outputs** - number of employees, total assets and the size of capital employed were chosen as inputs for evaluating the performance of companies. Capital employed represents equity, issued long-term bonds and bank loans and financial accommodations. Output is the economic value added (EVA), which, however, can reach negative values. When analyzing DEA the condition is that all input and output values are non-negative, therefore, REVA, an auxiliary indicator (5), was created. For companies with a positive EVA value, REVA is greater than 1. For companies with a negative EVA value, it is less than one.

$$REVA = 1 + \frac{EVA}{T} \tag{5}$$

EVA ... economic value added,
T ... production.

EVA was calculated in a modular way according to the methodology of the Ministry of Industry and Trade [11]. The calculation is based on equity and spread, i.e. the difference between return on equity and alternative cost of equity, see equation (6).

$$EVA = (ROE - r_e).E \tag{6}$$

$$ROE = \frac{PL}{E};\ r_e = \frac{WACC\,\frac{UZ}{A} - \frac{PL}{PLBT}IR\left(\frac{UZ}{A} - \frac{E}{A}\right)}{\frac{E}{A}};\ WACC = \frac{\frac{E}{A}r_{ef} + \frac{PL}{PLBT}IR\left(\frac{UZ}{A} - \frac{E}{A}\right)}{\frac{E}{A}}$$

$$r_{ef} = r_f + r_{pod} + r_{finstab} + r_{LA}$$

ROE ... return on equity,
r_e ... alternative cost of equity,
PL ... profit/loss of current accounting period,
E ... equity,
WACC ... weighted average cost of capital,
UZ ... capital employed,
A ... assets,
PLBT ... profit/loss before tax,
IR ... interest rate,
r_{ef} ... total risk premium,
r_f ... risk-free rate, fixed at a yield of ten-year government bonds, numerical values taken from MPO (2015),
r_{pod} ... business risk premium,
$r_{finstab}$... risk premium for financial stability which is linked to the L3 liquidity,
r_{LA} ... risk premium for the company size depending on the size of capital employed.

To eliminate price effects, the value data were expressed in constant prices of the year 2009. The figures in current prices were converted to the price level of 2009 using the industrial producer price index in the textile, clothing and leather industry [5].

3. **Determination of technical efficiency** – for each company in the cluster there were determined a score of technical efficiency E_q (assuming CRS - using a CCR model) and pure technical efficiency PE_q (assuming VRS, BCC model). In both cases, the models with a focus on inputs were used. Technical efficiency can be decomposed to pure technical efficiency and scale efficiency, see equation (4).

4. **Calculation of Malmquist index** - for each unit and period there was determined total factor productivity change and the change of its components (technological change, technical efficiency change and scale efficiency change). A model with the VRS assumption oriented at inputs was used. LibreOffice Calc 5.2.5 spreadsheet was used for data preparation and DEAP 2.1 software for calculations of DEA models [3]. According to the amount of values of technical and technological efficiency change, companies were divided into four quadrants – technological innovators, efficient – technologically backward, inefficient – innovative and inefficient and falling behind [14].

4 Research results

The average rate of pure technical efficiency (see Table 1) of companies in the cluster throughout the whole observed period of 2009-14 was not high. To achieve efficiency, the companies would have to reduce inputs by 62% to 81%. From a total of 17 companies, 2 to 5 companies were located on the efficient frontier. No company, however, lay on the efficient frontier throughout the whole observed period. Nevertheless, three companies (DMU 6, 9 and 10) were efficient in 5 periods out of a total of 6 observed. The development of company technical efficiency might have been influenced by the economic cycle when the Czech economy was in recession within the years 2009, 2012 and 2013. This connection, however, is difficult to prove scientifically as the overall economic situation of the economy is reflected in the financial results of companies with a time delay. Between the average level of technical efficiency and the rate of economic growth in the observed period was not found any statistically significant correlation (Spearman's rank correlation coeff. = -0.4286, p-value = 0.3379).

In the period of 2009-2014, Malmquist index showed relatively large fluctuations. After the 2010/09 annual decline, efficiency of companies in the textile cluster sharply increased in the following period. In the next two seasons it decreased and just before the end of the reporting period it increased.

DMU	Company	2009	2010	2011	2012	2013	2014
1	SINTEX	0.077	0.101	0.071	0.027	0.096	0.093
2	NYKLÍČEK	0.176	0.185	0.164	0.054	0.165	0.175
3	RETEX	0.063	0.064	0.042	0.011	0.056	0.054
4	ELAS	0.137	0.144	0.124	0.051	0.136	0.156
5	INOTEX	0.199	0.206	0.227	0.052	0.381	0.256
6	ARGUN	1.000	1.000	1.000	1.000	0.647	1.000
7	INTERCOLOR	0.068	0.072	0.019	0.026	0.096	0.095
8	S.P.M.	1.000	0.012	0.153	0.121	0.338	0.511
9	SILK & PROGRESS	1.000	1.000	1.000	1.000	1.000	0.328
10	ZITEX	1.000	1.000	1.000	0.379	1.000	1.000
11	STAP	0.045	0.052	0.065	0.014	0.045	0.068
12	VEBA	0.028	0.013	0.030	0.051	1.000	0.009
13	KOUTNÝ	0.357	0.193	0.359	0.023	0.597	0.138
14	SVITAP J.H.J.	0.018	0.020	0.017	0.008	0.022	0.026
15	ODETKA	0.223	0.234	0.171	0.091	0.256	0.367
16	NOVÁ MOSILANA	0.033	0.019	0.033	0.017	0.035	0.012
17	PAPILLONS	1.000	1.000	1.000	0.317	0.193	0.183
	AVG	**0.378**	**0.313**	**0.322**	**0.191**	**0.357**	**0.263**

Table 1 Pure technical efficiency scores

Year	Eq	Tq	PEq	SCq	Mq
2010/09	0.524	1.009	0.721	0.727	0.528
2011/10	0.883	4.590	1.128	0.783	4.054
2012/11	0.827	1.067	0.447	1.850	0.883
2013/12	2.607	0.240	3.156	0.826	0.625
2014/13	1.484	1.031	0.666	2.227	1.530
AVG	**1.082**	**1.041**	**0.948**	**1.141**	**1.126**

Table 2 Malmquist index summary of annual means

DMU	Eq	Tq	PEq	SCq	Mq
1	1.016	1.052	1.039	0.978	1.068
2	0.981	0.956	0.999	0.981	0.937
3	0.937	1.048	0.968	0.968	0.983
4	1.002	1.020	1.026	0.976	1.022
5	1.030	1.070	1.052	0.979	1.102
6	5.688	0.931	1.000	5.688	5.298
7	1.048	1.073	1.068	0.981	1.124
8	1.011	1.204	0.874	1.157	1.218
9	0.867	1.104	0.800	1.084	0.958
10	1.000	1.007	1.000	1.000	1.007
11	1.073	1.033	1.085	0.989	1.108
12	0.964	1.041	0.795	1.214	1.004
13	0.986	1.019	0.826	1.193	1.004
14	1.046	1.031	1.073	0.974	1.079
15	1.048	1.058	1.105	0.949	1.109
16	0.962	1.042	0.822	1.171	1.003
17	0.700	1.027	0.712	0.983	0.719
AVG	**1.082**	**1.041**	**0.948**	**1.141**	**1.126**

Table 3 Malmquist index summary of firm means

On the one hand, the causes of this phenomenon can be seen in the overall economic development and on the other hand, in changing the production characteristics of individual companies in the cluster. In 2009, 2012 and 2013 Czech economy showed economic downturn. By contrast, in 2010, 2011 and 2014 it saw economic growth [6]. Therefore, it can be stated that the development of Malmquist index follows the development of the Czech economy, given the length of time series, even if correlation coefficient is not statistically significant at the 5% level (Spearman's Rank Correlation Coeff. = 0.2319, p-value = 0.6041). Regarding the production characteristics of individual companies, major changes were recorded in ARGUN (discontinuity in number of employees from 1 to 17, assets from CZK 3.5 mil. to CZK 74 mil. and a capital from CZK 0.5 mil. to CZK 27 mil. between the years 2012 and 2013) and partly at SPM (decrease in number of employees between the years 2009-14 by 46%

and in assets by 70%), SVITAP (decrease in number of employees by 29%) and PAPILLONS (increase in number of employees by 76%, assets by 270% and capital by 648%).

Generally, it can be stated that on average efficiency of the companies investigated in the period of 2009-14 increased by about 13 % annually. Out of this, efficient frontier units increased their productivity by an average of 4 % and internal relative efficiency of companies with respect to efficient frontier units improved by an average of 8 %. Change in technical efficiency was driven in particular by improving scale efficiency, by an average of 14%. Conversely, pure technical efficiency of companies declined by about 5% (see Table 3).

5 Conclusion

We can conclude that efficiency of the companies in the cluster over the monitored period of 2009-2014 increased. The best companies are those that achieve efficiency improvements in both components. There were 9 such technological innovators in the monitored period (DMU 1, 4, 5, 7, 8, 10, 11, 14, and 15). Only one company (DMU 2) was inefficient with non-innovative technology. Other units can either be classified as efficient, but little innovative (one firm - DMU 6), or belong to the group of innovative but inefficient companies (6 companies - DMU 3, 9, 12, 13, 16 and 17). In terms of expected effects of the cluster on innovative performance and development of the member companies, the vast majority (15) of the companies confirmed the shift in technology in terms of efficient frontier changes. In the case of 10 companies technical efficiency improved, however, that was due to economies of scale. Especially in the case of small and medium-sized companies, it can be assumed that economies of scale occurred due to the inclusion of companies in the cluster. Conversely, organizational efficiency related to its operational management deteriorated. However, it is necessary to consider the fact that in the monitored period, companies had to contend with the consequences of the global economic crisis. Further research will focus on monitoring efficiency of the companies in the cluster in the coming years to extend the time series. Another promising area of research is monitoring the development of efficiency of companies within the same industry, but outside the cluster. This would provide the possibility to differentiate the real impact of the cluster on innovation and performance of member companies more precisely from development which these companies would achieve in the event that they were not members of the cluster.

References

[1] Bisnode: *Magnusweb: Komplexní informace o firmách v ČR a SR* [online]. Bisnode ČR, Prague, 2016 [cit. 2017-01-05]. Available at: https://magnusweb.bisnode.cz.

[2] Brožová, H., and Boháčková, I.: Post-crisis development in the agricultural sector in the CR – Evidence by Malmquist index and its decomposition. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics 2016* (Kocourek, A., and Vavroušek, M., eds.). Technical University of Liberec, Liberec, 2016, 73–78.

[3] Coelli, T.J.: *A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program.* University of New England, Armidale, 1996.

[4] Coelli, T. J., Prasada Rao, D. S., O´Donell, Ch. J., and Battese, G.: *An Introduction to Efficiency and Production Analysis*. Springer Science and Business Media, New York, 2005.

[5] CZSO: *Veřejná databáze – Indexy cen průmyslových výrobců* [online]. Czech Statistical Office, Prague, 2017 [cit. 2017-01-06]. Available at: https://vdb.czso.cz.

[6] CZSO: Veřejná databáze – Hrubý domácí produkt ČR výrobní metodou [online]. Czech Statistical Office, Prague, 2017 [cit. 2017-01-31]. Available at: https://vdb.czso.cz.

[7] Ferreira, D.C., Marques, R.C., and Pedro, M.I.: Comparing efficiency of holding business model and individual management model of airports. *Journal of Air Transport Management* **57** (2016), 168–183.

[8] Gonzales, E., Arrondo, R., and Carcaba, A.: Product innovation in the Spanish auto market: Frontier shift and catching-up effects. *Transportation Research Part D: Transport and Environment* **50** (2017), 170–181.

[9] Jablonský, J., and Dlouhý, M.: Modely hodnocení efektivnosti produkčních jednotek. Professional Publishing, Prague, 2004.

[10] Jiang, B., and Li, C.: Application of Nonparametric Model to Measure Productivity of Dairy Industry Cluster. *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia* **39** (2016), 80–86.

[11] MPO: *Finanční analýza podnikové sféry za rok 2014.* Ministry of Industry and Trade, Prague, 2015.

[12] Porter, M. E.: Clusters and the new economics of competition. *Harward Business Review* **76** (1998), 78.

[13] Řepková, I.: Efficiency Change of the Czech Banking Sector. In: *Proceedings of the International Conference Hradec Economic Days 2014, Part III.* (Jedlička, P., ed.). Gaudeamus, Hradec Králové, 2014, 48–54.

[14] Trueblood, M., and Coggins, J.: *Intercountry agricultural efficiency and productivity: A Malmquist index approach*. University of Minnesota, St. Paul, 2003.

[15] Zhu, J.: *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets.* Springer, Cham, 2014.

All papers passed a blind review process.

This publication has not been a subject of language check.

Published with the financial support of Czech Software First, Ltd., K Západi 54, 621 00 Brno.

Titlle: 35th International Conference Mathematical Methods In Economics, Conference Proceedings
Author: composite authors participants of the international conference
Intended: for participants of the internationalconference form researchers and professional public
Publisher: Gaudeamus, University of Hradec Králové
Published: in September 2017
Pages: 896
Edition: 1.
Press: Karo Reklama s.r.o.
Reference number of publication: 1615

ISBN 978-80-7435-678-0

University of Hradec Králové
Faculty of Informatics and Management

ČSOV

35th International Conference
Mathematical Methods in Economics

September 13th - 15th, 2017, Hradec Králové, Czech Republic

ISBN 978-80-7435-678-0