36th International Conference

Mathematical Methods in Economics

September 12th – 14th, 2018, Jindřichův Hradec, Czech Republic

Conference Proceedings

Published by:
MatfyzPress,
Publishing House of the Faculty of Mathematics and Physics Charles University
Sokolovská 83, 186 75 Praha 8, Czech Republic
as the 565. publication.
Printed by Reprostředisko MFF UK.

Printed in Prague — September 2018

Organized by:
Faculty of Management, University of Economics, Prague

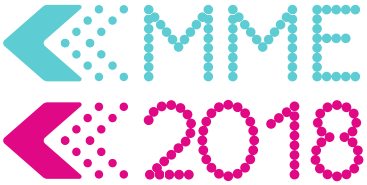
Under auspices of:
Czech Society for Operations Research
Czech Econometric Society

Credits:
Editors: Lucie Váchová, Václav Kratochvíl
LaTeXeditor: Václav Kratochvíl
Cover design: Jiří Přibil
using LaTeX's 'confproc' package, version 0.8 by V. Verfaille

ISBN: 978-80-7378-371-6 (printed version)
978-80-7378-372-3 (electronic version)

36th International Conference

Mathematical Methods in Economics



Organized by:
Faculty of Management,
University of Economics, Prague

Jindřichův Hradec

September 12-14, 2018

Programme Committee:

Miroslav Plevný - **chair**, *University of West Bohemia, Faculty of Economics*
Vladislav Bína, *University of Economics, Prague, Faculty of Management*
Ivan Brezina, *University of Economics Bratislava, Faculty of Economic Informatics*
Helena Brožová, *University of Life Science in Prague, Faculty of Economics and Management*
Martin Dlouhý, *University of Economics, Prague, Faculty of Informatics and Statistics*
Jan Fábry, *University of Economics, Prague, Faculty of Informatics and Statistics*
Petr Fiala, *University of Economics, Prague, Faculty of Informatics and Statistics*
Jana Hančlová, *Technical University of Ostrava, Faculty of Economics*
Josef Jablonský, *University of Economics, Prague, Faculty of Informatics and Statistics*
Radim Jiroušek, *University of Economics, Prague, Faculty of Management*
Miloš Kopa, *Charles University in Prague, Faculty of Mathematics and Physics*
Ladislav Lukáš, *University of West Bohemia, Faculty of Economics*
Michal Novák, *University of Economics, Prague, Faculty of Management*
Juraj Pekár, *University of Economics Bratislava, Faculty of Economic Informatics*
Jan Pelikán, *University of Economics, Prague, Faculty of Informatics and Statistics*
Jiří Přibil, *University of Economics, Prague, Faculty of Management*
Vladimír Přibyl, *University of Economics, Prague, Faculty of Management*
Jaroslav Ramík, *Silesian University in Opava, School of Business Administration in Karviná*
Karel Sladký, *Institute of Information Theory and Automation, Czech Academy of Sciences*
Tomáš Šubrt, *University of Life Science in Prague, Faculty of Economics and Management*
Jana Talašová, *Palacký University in Olomouc, Faculty of Science*
Milan Vlach, *Charles University in Prague, Faculty of Mathematics and Physics*
The Kyoto College of Graduate Studies for Informatics
Karel Zimmermann, *Charles University in Prague, Faculty of Mathematics and Physics*
Miroslav Žižka, *Technical University of Liberec, Faculty of Economics*

Foreword

In the years of its existence, the International Conference Mathematical Methods in Economics (MME), held especially under the auspices of the Czech Society for Operations Research, has established a stable position in its field. This year we can thus celebrate its 36^{th} annual meeting.

We would like to thank the authors of all submissions to the Mathematical Methods in Economics 2018. The Proceedings contain 112 contributions selected by the Programme Committee for presentation at this year's Conference. The papers cover various topics, such as Operations Research, Decision-Making, Mathematical Modeling, and Econometrics. We would like to extend our gratitude to the Programme Committee, not only for the reviews but also for their valuable comments and opinions. And, of course, we would like to thank all the external reviewers who contributed to the Conference with their evaluations.

After 19 years, the MME Conference is coming back to the beautiful small city of Jindřichův Hradec - the 17^{th} meeting was organized there on the September 14-16, 1999. The first references to Jindřichův Hradec date back to the 9^{th} century and are connected to the existence of a border fort. But the oldest preserved written report dates to 1220 - around that time, a Czech noble, Jindřich of Hradec, built a Gothic castle in the place of the former fort. Today, the complex of the old and new castles is the third largest in the Czech Republic (after Prague and Český Krumlov Castles). Together with the famous Krýza's Creche (the largest mechanical nativity scene in the world) and the narrow-gauge railway, they attract the attention of all visitors to the South Bohemia Region.

Last year, we celebrated the 25^{th} anniversary of the beginnings of university education in the town of Jindřichův Hradec. The MME 2018 Conference presentations will be given in lecture rooms of the Faculty of Management, University of Economics, Prague, on the Faculty's premises in Jindřichův Hradec. During the years of its existence, the Faculty of Management has become an integrated part of Jindřichův Hradec; today it provides the managerial education in the pleasant environment of a reconstructed modern building.

I wish all the participants a successful Conference and a pleasant stay in Jindřichův Hradec during the Mathematical Methods in Economics 2018.

In Jindřichův Hradec, August 2018 *Lucie Váchová*

List of Content

655 List of Authors

Client Interest Rates and Household Behavior: the Case of Czech Republic

Ondřej Badura[1]

Abstract. This paper deals with the question of causal relationship of interest rates and consumption-saving behavior of households. Although this issue has been at the forefront of empirical research for decades, yet many aspects of this relationship still remain the source of a broad polemic. One of the most questionable is the true nature of transmission from interest rates to tendency to saving. Above all due to the contradictory impacts of the substitution and income effect, the empirical results of classical approaches using single monetary-policy interest rate have been ambiguous, not very robust and thus not quite stable. However, the actual problem doesn't have to lie in the unclear nature of the relationship itself, but rather in the way we examine it. The aim of this study is to test a new method of verifying this relationship by dividing households into income categories according to their position in distribution of disposable income combined with the use of different types of client interest rates doing so on the example of the Czech Republic. As indicated by the initial results achieved through panel GMM regression, this new approach could become a relevant contribution to our understanding of transmission from interest rates to distribution of disposable income and thus to creation of aggregate demand.

Keywords: Client interest rates, saving rate, income categories, panel GMM.

JEL Classification: D12, D14
AMS Classification: 62P20

1 Introduction

Household consumption and savings are key indicators of economic activity both from the perspective of individuals, income groups, or at aggregate level. If the satisfaction of needs through consumption, as it is known, is the primary and final purpose of every economic effort, deciding between current consumption and current making of savings is de facto nothing else than a choice between meeting needs now and in the future. Since work of Friedman [10], interest rate is supposed to be one of the structural determinants of this household decision making. But does this assumption correspond to the economics reality? Although this question has been at the forefront of empirical research for decades, the answer still remains rather ambiguous.

When focusing particularly on savings, causal relationship of interest rate and savings rate is, at first glance, quite logical and relatively intuitive. Higher interest rate promising greater appreciation motivates households to save more. Then we should observe the proportional relationship between the interest rate and the propensity to save. Is this simple premise, however, in line with reality? But perhaps we should first ask whether this assumption is fully consistent with the current state of theoretical knowledge. Already the theory of intertemporal substitution in consumption has shown that reality is not that simple. As it is usual in the microeconomic analysis, we need to distinguish between the influence of substitution and income effect. In terms of current consumption, the interest rate actually reflects the opportunity costs. Together with the growth of potential money appreciation, consumption becomes relatively more expensive, on the contrary, making of savings more advantageous. So when interest rate rises, households are motivated to replace current consumption expenditures with savings. Thus the substitution effect affects the savings rate positively. The income effect, on the other hand, implies that higher interest also means a higher expected income and therefore the higher consumption both in the future and in the current period. The observed impact is therefore negative. In other words, the substitution effect says that a higher interest rate will make savings more attractive but also less necessary due to the income effect.

When looking for the description of this relationship in the standard consumption-savings theory, we find that the evidence is at best unclear. Life-Cycle Permanent Income Hypothesis that was for a long time the dominant approach in analyzing budgetary behavior of households, is primarily based on works of Modigliani and Brumberg [14] and Friedman [10]. Although these works are in principal very similar, they significantly differ in one thing after all. Friedman [10], unlike Modigliani and Brumberg [14], directly defines the marginal propensity to consume as a function of the real interest rate, but, unfortunately, he does not comment on this relationship

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

in more detail. However, according to the theory of permanent income the propensity to consume (and propensity to save) is influenced by (permanent) income expectations rather than by interest rate itself. In the case of Friedman [10] adaptive, later also rational expectations, as pointed out by Hall [11].

As demonstrated later by Campbell and Mankiw [6], when deciding about the amount of consumption and savings, only a part of the population take into account the permanent income, and therefore the functional relationship between the interest rate and the propensity to consume, as assumed by Friedman [10], would be significantly weakened. Flavin [9] offers one of the most promising explanations for this phenomenon. She points to the fact that not all members of society (especially youth, socially disadvantaged groups, etc.) have a full-fledged approach to financial markets. They can't borrow in order to fully cover the expenses that would correspond to their permanent income. Hence, the interest rate, at which these additional finances might be provided, is not that important for them in their consumption decision-making. These perceptible liquidity constraints, together with a kind of myopia, i.e. the fact that not everyone is forward-looking, significantly weaken the position of the LC-PIH concept, and the relationship of interest rate and consumer behavior remains unclear. Instead, the impact of the interest rate on savings creation should be strengthened if the theory of Carroll [7] is valid. It assumes that people target the amount of their financial reserves. If this is the case, then the interest rate should be one of the key determinants in deciding on the creation of these savings, where, in addition to the substitution, the income effect would also take on the strength.

There are also a plenty of empirical studies dealing with the unclear nature of the interest rate – savings relationship. The predominance of income effect could be found for instance in panel data analysis of Loayza, Schmidt-Hebbel and Servén [12] or Mody, Ohnsorge and Sandri [15] revealing the significant precautionary effect after global economic crisis. Conversely, the superiority of the substitution effect can be found in the work of Masson, Bayoumi and Samia [13] or Aizenman, Cheung and Ito [1] who best describes the confusion and ambiguity of the empirical results in this area when on the one hand confirming the predominance of a positive effect between the trends in interest rate and propensity to save, but at the same time they point out that at very low interest rate levels the effect may also be the opposite. As demonstrated by aforementioned examples (and many other studies) the empirical results so far are quite contradictory. Where one research reveals the positive impact of the interest rate on the propensity to save, the study will eventually come with the opposite claim. The particular form of observed dependence is probably not a matter of general validity, but it's very sensitive to the specific characteristics of the given economy, time and, of course, the setting of the model used.

The aim of this study is to make a little contribution to this empirical puzzle by testing a new method of verifying this relationship by dividing households into income categories according to their position in distribution of disposable income combined with the use of different types of client interest rates doing so on the example of the Czech Republic.

To achieve this goal we use econometric analysis of dynamic panel data. However, our analysis will differ significantly in two points from most of the contemporary works with a similar focus. First, we don't take into account a single monetary interest rate, but a variation of client interest rates, as already mentioned in the aim of the paper. Secondly, we don't examine an impact of the interest rate on one aggregate propensity to save, but on the tendency to save of individual income groups. This approach can potentially open a new way of analyzing of interest rate pass-through in terms of households behavior.

The paper is organized as follows. In the second section we introduce the data, the model and the estimation method. The results of this analysis are presented in the section number 3. In the fourth section we conclude.

2 Data, methods and model

2.1 Data

We test the causal relationship of interest rate and savings rate using the example of the Czech Republic. The values of client interest rates were taken from the database of Czech National Bank, all other data from the Czech Statistical Office. In all cases, they are annual observations between 2004 and 2015, where only client interest rates have been recalculated as the simple average of individual monthly rates.

A key indicator of a proper understanding of household income and consumption expenditure data is their primary breakdown into ten sub-categories (panels) by deciles ordered ascendingly according to disposable income. The default time series of average consumption and disposable income per capita, expressed nominally[2] in CZK, were then used to calculate the average propensity to save as a simple share of savings on the disposable

[2] Although we can generally expect the impact of the interest rate on the real variables, due to the relative nature of *APS* and Y_{RD} indicators, the effect of changes in the price level is to be completely eliminated anyway.

income. The main explanatory variable - the relative disposable income in a sense of Duesenberry [8] was derived particularly in a definition used by Badura [4]:

$$Y_{RD_{i,t}} = \frac{Y_{D_{i,t}}}{\bar{Y}_{D_t}} \tag{1}$$

where the relative disposable income of the *i*-th income category at time *t* ($Y_{RDi,t}$) is calculated as the share of the disposable income of the same category $Y_{Di,t}$ on its whole-society weighted average Y_{Dt} that is formulated as the weighted average of disposable incomes of all categories in given time, particularly by the pattern:

$$\bar{Y}_{D_t} = \frac{\sum_{i=1}^{N} Y_{D_{i,t}} w_{i,t}}{\sum_{i=1}^{N} w_{i,t}} \tag{2}$$

where weights are set as the average numbers of household members in a given income category ($w_{i,t}$) and where *N* is the number of these categories.

There is a list of variables that we use in the following analysis: *APS* – average propensity to save, Y_{RD} – relative disposable income, *inf* – inflation rate, *u* – general unemployment rate, *age* – average age of the population, *ratio* – ratio of old-age pension to average wage, *balance* – share of general government balance to nominal GDP, *i1* – nominal[3] interest rate on overnight deposits, *i2* – nominal interest rate on deposits redeemable at notice, *i3* – nominal interest rate on consumer credit, *i4* – nominal interest rate on mortgages, *i5* – nominal interest rate on overdraft loans, *i6* – nominal interest rate on credit card receivables.

2.2 Methods and model

To achieve the objective stated above we use primarily econometric analysis. Compared to most studies exploring this topic (see introduction), however, our empirical analysis fundamentally differs in two areas. First, we don't examine the impact of interest rates on a single aggregate savings rate, but on the propensities to save of individual income categories. Unlike studies that do the contrary, we will be able to take into account the heterogeneity with which the different income groups react on basic determinants of their budgetary decision making.

The second major change is that we don't use a single monetary interest rate, but more types of client interest rates together. Why this change? The first reason is the simple premise that when examining the determinants of the household decisions about the distribution of disposable income, it is far more appropriate for its analysis to use the variables with which the households actually come into contact. In other words, if a study finds a significant causal relationship between the monetary-policy interest rate and the savings rate, it must necessarily be a secondary or mediated relationship. It is not possible for households to be directly affected by the type of interest rate they will never come into contact with. If we want to dive into the very essence of the problem and examine the primary influence of these two variables, it is necessary to use the interest rate at which households can actually borrow or save, that is, a client interest rate. The second advantage of this procedure is also the fact that we do not have to rely on a single variable, but we can use a wide range of client rates that themselves can work differently. Both the income and the substitution effect can, of course, prevail both in interest rates on deposits and on loans.

As the structure of the dataset already suggests, a panel regression is used to estimate the wanted equation, where individual entities are represented by the income categories of the population. As Hall [11] first demonstrated in his "random walk" model and as later demonstrated by countless empirical studies, when analyzing consumption-savings behavior over time, it is always necessary to take into account a certain persistence, therefore, it is not possible to use the static estimation of random or fixed effects, but the dynamic model needs to be estimated. The default estimated regression model M1 then takes the following form:

$$APS_{i,t} = \beta_0 \cdot APS_{i,t-1} + \beta_1 \cdot X_t + \beta_2 \cdot Z_{i,t} + \gamma + \mu_i + \varepsilon_{i,t} \tag{3}$$

where the average propensity to save $APS_{i,t}$ of the *i*-th income category at time *t* is estimated using its lagged value $APS_{i,t-1}$, the vector of variables of our interest (client interest rates) X_t and the vector of control variables $Z_{i,t}$. β_0 is a scalar expressing the sensitivity of average propensity to save on its lagged values, β_1 is the vector of

[3] As can be seen, nominal rates are used in all cases, although according to economic logic, primarily the real interest rate should have the influence on the households' budgetary behavior. However, we don't lose any information in this case as the inflation rate is also included as a separate indicator. On the contrary, by the elementary decomposition of real interest rate, we can easily separate the influence of nominal variables and movements in the price level. Households can no doubt identify these separate indicators much better than the resulting real interest rate.

the estimated coefficients for the client interest rates and β_2 the vector of estimations for the controls. γ denotes the level constant, μ_i is the unobserved individual effect and $\varepsilon_{i,t}$ then the remaining residual component.

The vector X_t is defined by dimensions 6×1 and it includes all client interest rate variables. The vector $Z_{i,t}$ is 6×1 as well and it includes variables: $Y_{RD,}$ *inf, u, age, ratio.* For the purpose of robustness check we introduce also alternative model M2, where vector $Z_{i,t}$ above that includes also variable of *balance*. All control variables were chosen both on the basis of the theoretical knowledge about consumption-savings function, mainly from the postulates of Life Cycle Hypothesis (indicators of *age, ratio*) and modern empirical works (see the introduction). At the same time, however, it was necessary to take into account the sub-aggregated nature of the data, especially by implementation of relative income in definition by Badura [4].

Vectors of estimated parameters β_1 and β_2 both take dimension 1×6, where β_1 includes the parameters of our main interest. Because both substitution and income effect can prevail in relationship of savings rate and interest rate, as many studies have confirmed, neither the positive or negative values of the coefficients of vector β_1 can be explicitly assumed, they may even reach positive or negative values depending on the specific interest rate. Therefore, the signs of wanted coefficients are by no means the premise of this analysis, but its result, which will be discussed in the final chapter.

Due to the dynamic nature of the model and the fact that also many explanatory variables in the model may not only be exogenous, the most appropriate method of estimation seems to be a general approach of dealing the endogeneity issue using instrumental variables that was introduced by Arellano and Bond [2], i.e. FD_GMM (First-Difference Generalized Method of Moments) with the coefficient estimator pro the given *x* variable:

$$\hat{\beta} = \left[(\Delta x_{-1})' W (W' (I_N \otimes G) W)^{-1} W' (\Delta x_{-1}) \right]^{-1} \times \left[(\Delta x_{-1})' W (W' (I_N \otimes G) W)^{-1} W' \Delta x \right] \tag{4}$$

where *W* stands for the matrix of instruments of lagged levels of *x* given by the matrix *G*, which by definition is given by T-2×T-2. We also use the approach latter derived by Arellano and Bover [3], and Blundell and Bond [5], now called as S_GMM (System Generalized Method of Moments), which provides another additional instruments using the estimator for the given *x* variable:

$$\hat{\beta} = \left[\sum_i (W_i \otimes m_i)' \left(\hat{\Omega} \otimes \sum_i m_i m_i' \right)^{-1} \sum_i (W_i \otimes m_i) \right]^{-1} \sum_i (W_i \otimes m_i)' \times \left(\hat{\Omega} \otimes \sum_i m_i m_i' \right)^{-1} \sum_i (x_i \otimes m_i) \tag{5}$$

where Ω is a covariance matrix, m_i represents the subset of explanatory variables uncorrelated with random error and it stands as the last element of diagonal matrix of instruments W_i that is now given by moment conditions both in levels and in differences. For the purpose of robustness check and possible comparison of results we use both FD-GMM and S-GMM, where only variable *age* is assumed to be exogenous and all other variables are first treated endogenous and then the model is estimated again assuming those variables as predetermined.

3 Results

Although the input time series aren't too long, it is useful to check for the stationarity first. It was done so by 3 tests based on the null hypothesis of the unit root existence: Harris and Tzavalis, Breitung and Im, Pesaran and Shin[4].As the condition of stationarity in our dataset due to the length of time series is not so strict, we consider stationary only those variables for which the zero hypothesis was rejected by all tests at significance level of 0.1 at least. In this case, our requirements are met by indicators of *APS, i3, i6, YRD, inf, ratio, balance*. The other variables were converted to differences (suffix "*_d*") in order to remove the unit root and tested again. Our stationarity condition is then met by all modified variables but *i2_d*. Let's note that, when converting to the second difference, this variable finally fulfills the condition of stationarity, but this time at too high cost. It should be borne in mind that with each differentiation we lose N (number of entities) observations and due to a smaller number of input time periods, another such loss for the sake of one variable is not acceptable. For this reason, the variable *i_2* is completely discarded from both M1 and M2 models, and the indicators of *u, age, i1, i4, i5* are replaced by their differences.

The next step in procedure of data preparing for the regression estimation was to verify whether there is a significant multicollinearity among the independent variables. For this purpose, we use the matrix of Pearson correlation coefficient values. While it is worth keeping in mind at this point that there is no natural boundary separating the multicollinearity from an acceptable correlation rate, the value of the correlation of 0.8 is very

[4] The results couldn't be displayed due to a limited length of the article. The same applies also to the following test for multicollinearity and Arellano-Bond and Sargan-Hansen tests for the final model estimates.

often considered to be the limit.[5] Following this rule, control variables of *inf* and *u_d* had to be removed both from the M1 and M2 models as they are highly correlated with client interest rates (which are far more essential for our analysis, so they had to stay).

The resulting estimates and their p-values (in brackets) for all types of models are shown in Tables 1 and 2. For the purpose of robustness check, both model M1 and M2 are estimated using FD_GMM and S-GMM, assuming explanatory variables endogenous (suffix "-e") and predetermined (suffix "-p"). In accordance with the work of Arellano and Bond [2] a robust estimation of the standard error was used to ensure the residual homoscedasticity condition. We also need to note that all variants of the estimated models passed the Sargan-Hansen test without any problem, which indicates the correct specification of the model due to the exogenous nature of all instruments used. Using Arellano-Bond test we also fail to reject the null hypothesis of no second-order serial correlation for residual differences for any of the reflected models.

	M1 (FD-GMM-e)	**M1 (FD-GMM-p)**	**M1 (S-GMM-e)**	**M1 (S-GMM-p)**
i1_d (overnight deposits)	0.056 (0.123)	0.056 (0.127)	0.063 (0.104)	0.062 (0.105)
i3 (consumer credit)	0.017 (0.000)	0.017 (0.000)	0.018 (0.000)	0.017 (0.000)
i4_d (mortgages)	-0.015 (0.018)	-0.015 (0.018)	-0.015 (0.020)	-0.015 (0.020)
i5_d (overdraft loans)	0.034 (0.000)	0.034 (0.000)	0.033 (0.000)	0.033 (0.000)
i6 (credit cards)	-0.03 (0.000)	-0.03 (0.000)	-0.029 (0.000)	-0.029 (0.000)

Table 1 Estimation results for model M1

	M2 (FD-GMM-e)	**M2 (FD-GMM-p)**	**M2 (S-GMM-e)**	**M2 (S-GMM-p)**
i1_d (overnight deposits)	0.057 (0.114)	0.056 (0.119)	0.064 (0.093)	0.062 (0.096)
i3 (consumer credit)	0.016 (0.000)	0.016 (0.000)	0.017 (0.000)	0.017 (0.000)
i4_d (mortgages)	-0.014 (0.113)	-0.014 (0.117)	-0.014 (0.127)	-0.014 (0.123)
i5_d (overdraft loans)	0.034 (0.000)	0.034 (0.000)	0.033 (0.000)	0.033 (0.000)
i6 (credit cards)	-0.031 (0.000)	-0.031 (0.000)	-0.03 (0.000)	-0.03 (0.000)

Table 2 Estimation results for model M2

Due to a limited length of the article, we focus only on the variables of our interest and not on all other controls. First of all we need to emphasize that estimates for all client interest rates are very stable, their signs, magnitude and p-values (except for *i4_d)* do not change significantly across all types of models and methods. Therefore, we can proclaim achieved results as final and relatively reliable.

When evaluating the results for particular interest rates, we can say right from the start that the influence of interest rate on overnight deposits is insignificant. However, due to very low rates in this financial product, households aren't likely to be influenced by them and so this result could barely be called surprising. The situation is much more complicated when looking at result for interest rate on mortgages. In model M1 this coefficient appeared to be significant while the p-values in M2 do not allow us to claim the same. Although the very influence of this variable is unclear, we can say that if households are truly influenced by this type of interest rate, the income effect prevails, which implies that the higher interest rate doesn't dissuade Czech households from mortgage loans too much, but the higher costs in a form of elevated rates are absorbed into households' budget which is reflected in a lower propensity to save.

The influence of interest rates on consumer credit is no doubt statistically significant. The substitution effect is obviously stronger than the income effect in this relationship, which only reflects the fact that Czech households respond to higher rates with deferred consumption and thus with a lower demand for this banking product rather than with lowering their propensity to save. Finally, significant and relatively stable coefficients of interest rates from overdraft loans and credit cards have appeared to be slight surprise, since that is not, where would seek for this causality too much. What's more, although, these products are relatively close substitutes, according to results their effect on the average propensity to save is exactly the opposite. While the substitution effect dominates for the overdraft loans, the influence of interest rates for credit cards is negative and rather it reflects the classical position of households in the role of the borrower.

[5] Although this method is rather a rule of thumb then an accurate evaluation of multicollinearity, for smaller sample size like this it can be considered sufficient.

4 Conclusion

The aim of this work was to verify the causal relationship of client interest rates and the average propensity to save on the example of Czech households. For 4 out of 5 interest rates that we use, the above-mentioned relationship has been actually verified. In two cases the substitution effect prevailed, for the other two rates it was income effect the dominant one. It is so obvious that in the environment of the Czech Republic the influence of a single monetary interest rate would be ambiguous and probably also relatively weak. Therefore it is not possible to a priori assume the negative and positive impact of the interest rate on the propensity to save without knowing a particular type of the interest rate.

There was a relatively large innovation compared to studies investigating a similar topic that was the distribution of households into groups according to their position in the distribution of disposable income. However, the use of a set of client interest rates instead of one monetary policy rate has finally turned out to be even a greater benefit. Exactly as expected, it has been proven that economic units respond differently to different types of client rates. Once a negative relationship can prevail, in case of a different type of rate households can respond positively in their tendencies to save. This fact is a very likely an explanation of the current contradictory results when using a single monetary interest rate, which impact on the savings rate is actually determined by a set of effects from individual client rates. One of the major conclusions that is implied by the results of this study is that, unlike most of the current empirical research, the ambiguous and unstable relationship between monetary policy interest rate and propensity to save at aggregate level is not a surprise and source of confusion but a logical and inevitable consequence of heterogeneity across effects of client interest rates. Our approach thus potentially opens up a new way of analyzing the interest pass-through in relation to households.

Acknowledgement

This paper was supported within the VŠB–Technical University SGS grant project No. SP2018/138.

References

[1] Aizenman, J., Cheung,Y. W. and Ito, H. (2016). The Interest Rate Effect on Private Saving: Alternative Perspectives. NBER, *Working Paper No. 22872*.

[2] Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2), 277–297.

[3] Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29–51.

[4] Badura, O. (2016). An Influence of Relative Income on the Marginal Propensity to Consume – Evidence from Czech Republic. In Majerová, I. and Kotlánová, E. (Eds.), *Proceedings of 14th International Scientific Conference: Economic Policy in the European Union Member Countries* (pp. 21-33). Petrovice u Karviné: Department of Economics and Public Administration of the Silesian University in Opava.

[5] Blundell, R. W. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.

[6] Campbell, J. Y. and Mankiw, G. N. (1989). Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence. *NBER, Macroeconomics Annual 1989*.

[7] Carroll, Ch. D. (1997). Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis. Quarterly *Journal of Economics*, 107, 1-56.

[8] Duesenberry, J. S. (1949). *Income, Saving, and the Theory of Consumer Behavior*. Cambridge: Harvard University Press.

[9] Flavin, M. A. (1985). Excess Sensitivity of Consumption to Current Income: Liquidity Constraints or Myopia? *Canadian Journal of Economics*, 18, 117-136.

[10] Friedman, M. (1957). *A Theory of the Consumption Function*. New York: Princeton University Press.

[11] Hall, R. E. (1978). Stochastic Implication of the Life Cycle – Permanent Income Hypothesis: Theory and Evidence. *Journal of Political Economy*, 86, 971-987.

[12] Loayza, N., Schmidt-Hebbel, K. and Servén, L. (2000). What Drives Private Saving across the World. *Review of Economics and Statistics*, 82(2), 165-181.

[13] Masson, P. R., Bayoumi, T. and Samiei, H. (1998). International evidence on the determinants of private saving. *The World Bank Economic Review*, 12(3), 483-501.

[14] Modigliani, F. and Brumberg, R. (1954). Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data. In Kurihara, K. (Eds.), *Post-Keynesian Economics* (pp. 338-436). New Brunswick: Rutgers University Press.

[15] Mody, A., Ohnsorge, F. and Sandri, D. (2012). Precautionary Saving in the Great Recession. *IMF Working Paper No. 12/42.*

Europe 2020 Climate Change and Energy Objectives in Central European Countries. Measurement via taxonomic measure of development with generalized distance measure GDM

Adam P. Balcerzak[1]

Abstract. The main objective of the article is to compare results obtained by Central European countries in regard to reaching climate change and energy efficiency objectives given in the Europe 2020 plan. The economies were chosen for the research, as they form relatively homogenous group in the sphere of socio-economic development and to relatively high extent their national energy systems are based on conventional energy resources. Based on the Europe 2020 plan there are five climate change and energy criteria, which should be applied for assessment of the economies. Thus, the problem should be analysed with application of multiple-criteria analysis methods. In the research taxonomic measure of development method, which was proposed by Hellwig, was used. The original Hellwig's method was based on the Euclidean metric as a distance measure. However, in the current research the applied method was modified and was based on the generalized distance measure GDM. The research was done for the years 2005, 2010 and 2015. The method enabled to propose rankings, which can be useful for benchmarking the economies and looking for examples of good practices in national energy policy.

Key words: climate change and energy, central European countries, Europe 2020 plan, Hellwig's method, generalized distance measure GDM, multiple-criteria analysis.

JEL Classification: C38, O13

AMS classification: 91C20

1 Introduction

Climate change and energy efficiency objectives have been influencing European economic policy for last two decades. They have been included both in the Lisbon Strategy and the Europe 2020 plan [34]. Therefore, they influence significantly national strategies concerning sustainable development and national energy security policies [36; 24]. The application of climate change and energy efficiency objectives is also supported by European funds, which justifies the need for monitoring and comparing European economies in this regards. Therefore, the objective of the current research is to compare achievements of Central European economies in the field of obtaining climate change and energy objectives given in the Europe 2020 plan.

The set of countries under research was limited to the region of Central Europe deliberately. From the perspective of socio-economic development the countries form relatively homogenous group. Due to historical common background all the countries are affected by the process of technological modernisation of their economies, which is especially difficult in the case of national energy systems, mostly based on conventional energy resources.

In the Europe 2020 plan five main targets relating to employment, innovation level, climate and energy utilisation, education and poverty were given. In the case of every target a set of specific measurable objectives was proposed. For climate change and energy target there are five specific criteria, which were pointed by European Commission and provided by Eurostat, that should be used in the process of assessment of the European economies. Therefore, the analysed phenomenon should be considered as multiple-criteria analysis problem. In the current research taxonomic measure of development (TMD) with generalized distance measure GDM was applied. The analysis was conducted for the years 2005, 2010, 2015.

2 Methodology

Most of economic problems should be considered as complex multiple-criteria phenomena, which cannot be assessed with application of one simple diagnostic variable [38; 3; 28; 33; 44; 43; 13; 35; 20; 21; 22]. This situation creates scientific need for application of taxonomy and multiple-criteria methods in empirical

[1] Adam P. Balcerzak, PhD, Nicolaus Copernicus University in Toruń, Faculty of Economic Sciences and Management, Ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: adam.balcerzak@umk.pl

economics [7; 11; 12; 19; 17; 4, 27; 29; 6; 37; 18]. It is especially important in the case of international comparative research or studies that should cover the problem of spatial dependencies [23; 30; 31; 39; 8; 40; 32; 14; 15; 16]. In order to apply multiple-criteria analysis method an economic phenomenon should be characterised with a set of aspects, where available diagnostic variable should enable to describe specific part of the phenomenon under research [26].

In the case of every application of multiple-criteria method the crucial role can be attributed to the process of selection of diagnostic variables. The selection process of diagnostic variables is usually affected by subjective factors related to the experience and specialist knowledge of researchers. Therefore, a set of potential diagnostic variables should be verified with application of formal criteria of high information value. The final set of diagnostic variables should be classified into stimulants and dis-stimulants and normalised.

In the case of current research the method of taxonomic measure of development (TMD) proposed by Hellwig [9; 10] was applied. The main advantage of the method, which is the source of its common applicability in multiple-criteria research in economics, is its methodological simplicity [1; 5; 25]. The TMD enables to describe all determinants of analysed economic phenomenon and it allows to obtain synthetic measure of its level, which can be used for ranking of objects. The analysed objects are compared to pattern of development, which is a set based on a maximum value of diagnostic variables for stimulants. In the case of dynamic research the pattern must be constant for the whole research period.

In the original TMD method proposed by Hellwig an Euclidean metric was used to estimate the distance of the object from the pattern of development. However, in the current research the measure GDM proposed by Walesiak [41] was applied. The metric was chosen due to its universality, as it can be used in the case of variables measured on the ratio scale, interval scale, the ordinal scale or the nominal scale. The values of generalized distance measure GDM from pattern of development (GDM_{it}^{P}) are assessed with equations 1[2].

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

$i, l = 1, \ldots, n$ – number of the object, k – number of pattern of development, $j = 1, \ldots, m$ – number of variable, Z_{ijt} – normalized diagnostic variable, P_j – pattern of development,.

The estimation can be conducted with Package ‘clusterSim” [42]. After obtaining the TMD with application of Hellwig procedure, it is possible to group the analysed objects into relatively homogenous subsets, which can be done with application of natural breaks method.

3 Results

The aim of the article is to compare results of Central European economies: Estonia, Latvia, Lithuania, Poland, Czech Republic, Slovakia, Hungary, Romania, Bulgaria, Slovenia in the process of reaching climate change and energy efficiency objectives in the years 2005-2015. The research period was limited with the availability of data for all indicators proposed by European Commission and provided by Eurostat. The set of five variables given for the international evaluation purposes with their classification is given in table 1.

Variable	Description of the variable	Variable Character
X_1	Greenhouse gas emissions, base year 1990 - Index (1990 = 100) - the indicator shows trends in total man-made emissions of the ‘Kyoto basket’ of greenhouse gases. It presents annual total emissions in relation to 1990 emissions	Dis-stimulant
X_2	Share of renewable energy in gross final energy consumption in %	Stimulant
X_3	Primary energy consumption - Million tonnes of oil equivalent (TOE) in relation to population - it is meant the Gross Inland Consumption excluding all non-energy use of energy carriers. This quantity is relevant for measuring the true energy consumption and for comparing it to the Europe 2020 targets.	Dis-stimulant

X_4	Final energy consumption - Million tonnes of oil equivalent (TOE) in relation to population - it is meant all energy supplied to industry, transport, households, services and agriculture. This quantity is relevant for measuring the energy consumption at final place of energy use and for comparing it to the Europe 2020 targets.	Dis-stimulant
X_5	Greenhouse gas emissions in ESD sectors - ESD base year=100 / million tonnes CO2 equivalent in relation to population - The indicator calculation is based on the emissions covered under the Effort Sharing Decision, which sets national annual binding targets for emissions not covered under the EU emission trading scheme (ETS). The ESD emissions are calculated by deducting ETS verified emissions, CO2 emissions from domestic aviation and NF3 emissions from national total emissions.	Dis-stimulant

Table 1 Diagnostic variables

In the research the Hellwig's method with application of generalized distance measure GDM, which was described in second section was used. After obtaining the TMD measure and proposing the rankings of the countries, they were grouped in three homogenous classes with application of natural breaks methods, where the class 3 is characterised with the "best" results and the class 1 can be considered as the "worst" one. The results are given in table 1 and figure 1.

Country	**2005**		**2010**		**2015**	
	TMD	**Class**	**TMD**	**Class**	**TMD**	**Class**
Romania	0.859	3	0.950	3	0.920	3
Bulgaria	0.678	3	0.812	3	0.827	3
Hungary	0.519	2	0.722	3	0.784	3
Lithuania	0.844	3	0.876	3	0.777	3
Slovakia	0.316	2	0.369	2	0.524	2
Latvia	0.397	2	0.387	2	0.387	2
Poland	0.465	2	0.358	2	0.384	2
Estonia	0.335	2	0.347	2	0.321	2
Slovenia	0.155	1	0.208	1	0.200	1
Czech Republic	0.069	1	0.121	1	0.035	1

Table 2 Ranking and grouping of the countries' in the years 2005, 2010, 2015

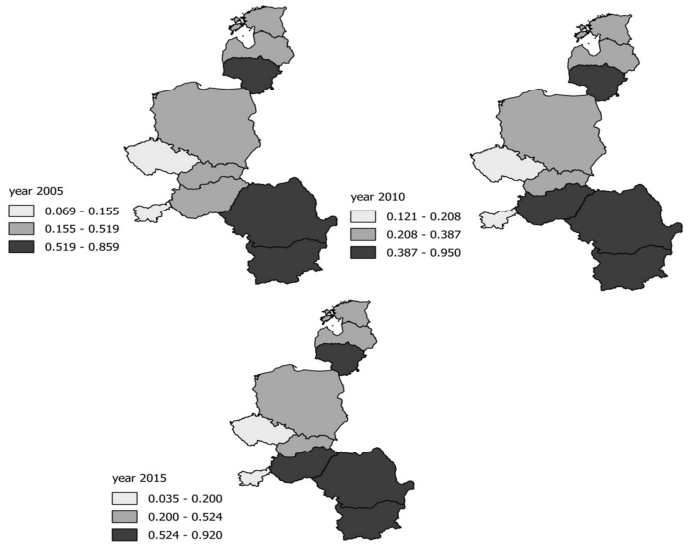


Figure 1 Grouping of the Central European countries with application of natural breaks method

In the subset grouping the countries obtaining the "best" results in all three analysed years one can find Romania, Bulgaria and Lithuania. Hungary joined this group in the years 2010 and 2015. In the second subset during the whole period there are Slovakia, Latvia Poland and Estonia. Slovenia and Czech Republic were rated as the once with the worst results in regard to the analysed phenomenon.

The analysis of the ranking in the whole period confirms high stability of the obtained results. As it was mentioned, only Hungary was able to improve its relative position and move to the "better" class. As a result Hungary can be considered as an example of a country that was able to combine relatively high level of standard of living based on high level of industrialisation with relatively good results in reaching the climate change and energy objectives. In this last context high positions of Romania and Bulgaria, two economies with the lowest GDP per capita in Central Europe, and the lowest position of Czech Republic, which is the best developed country in the region, can indicate significant difficulties in reaching economic welfare objectives with climate change and low emission principals.

4 Conclusions

The aim of the article was to compare the results of Central European economies in the field of climate change and energy efficiency objectives given in the Europe 2020 plan. In the research taxonomic measure of development method with application of measure GDM was used for proposing three rankings for the years 2005, 2010 and 2015. Based on the value of the measure the countries were grouped into homogenous subsets.

The main empirical contribution of the article can be attributed to two points. First of all, the research can be considered as an argument in favour of the view that there are important contradictions between the level of economic development and climate change and energy efficiency objectives. Second, the research confirm relatively high stability in time of the results obtained by given countries. This outcome confirms the role of long term factors that influence the energy consumption patterns of economies. From the perspective of policy implications it confirms that historical and structural determinants of national energy systems must be taken into consideration during the process of forming middle and long term sustainable development strategies at national and international level.

References

[1] Balcerzak, A. P. (2016). Multiple-criteria Evaluation of Quality of Human Capital in the European Union Countries. *Economics & Sociology*, 9(2), 11-27.

[2] Balcerzak, A. P., & Pietrzak, M. B. (2017). Digital Economy in Visegrad Coutnries. Multiple-criteria decision analysis at Regional Level in the Years 2012 and 2015. *Journal of Competitiveness*, 9(2), 5-18.

[3] Balcerzak, A. P., Kliestik, T., Streimikiene, D., & Smrčka L. (2017). Non-parametric approach to measuring the efficiency of banking sectors in European Union Countries. *Acta Polytechnica Hungarica*, 14(7), 51-70.

[4] Cano, J., Campo, E., & Gómez-Montoya, R. (2017). International Market Selection Using Fuzzy Weighing and Monte Carlo Simulation. *Polish Journal of Management Studies*, 16(2), 40-50.

[5] Cheba, K., & Szopik-Depczyńska, K. (2017). Multidimensional comparative analysis of the competitive capacity of the European Union countries and geographical regions. *Oeconomia Copernicana*, 8(4), 487-504

[6] Cyrek, M. (2017). Social efficiency of employment in three sectors — a comparison of Polish regions. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(3), 417-432.

[7] Ginevičius, R., Gedvilaitė, D., & Bruzgė, Š. (2015). Assessment of a Country's Regional Economic Development on the Basis of Estimation of a Single Process (ESP) Method. *Entrepreneurial Business and Economics Review*, 3(2), 141-153

[8] Hadaś-Dyduch, M. (2016). MODEL - Spatial Approach to Prediction of Minimum Wage In M. Reiff & P. Gezik (Eds.). *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XVIII*. Vratna: Letra Interactive, 134-138.

[9] Hellwig, Z. (1968). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przeglad Statystyczny*, 4, 307-327.

[10] Hellwig, Z. (1972). Procedure of Evaluating High-Level Manpower Data and Typology of Countries by Means of the Taxonomic Method. In Z. Gostowski (Ed.). *Towards a System of Human Resources Indicators for Less Developed Countries, Papers Prepared for a UNESCO Research Project.* Wrocław: Ossolineum-The Polish Academy of Sciences, 115-134.

[11] Keshavarz Ghorabaee, M., Zavadskas, E.K., Amiri, M., & Antucheviciene, J. (2016). Evaluation by an Area-based Method of Ranking Interval Type-2 Fuzzy Sets (EAMRIT-2F) for Multi-criteria Group Decision-making. *Transformations in Business & Economics*, 15(3) (39), 76-95.

[12] Keshavarz Ghorabaee, M., Amiri, M., Kazimieras Zavadskas, E. & Antucheviciene, j. (2017). Supplier evaluation and selection in fuzzy environments: a review of MADM approaches. *Economic Research-Ekonomska Istraživanja*, 30(1), 1073-1118.

[13] Kruk, H., & Waśniewska, A. (2017). Application of the Perkal method for assessing competitiveness of the countries of Central and Eastern Europe. *Oeconomia Copernicana*, 8(3), 337–352.

[14] Kuc, M. (2017b). Social convergence in Nordic countries at regional level. *Equilibrium-Quarterly Journal of Economics and Economic Policy*, 12(1), 25-41.

[15] Kuc, M. (2017c). Spatial differences in sustainable development components in Nordic regions. In M. Papież & S. Śmiech (Eds.). Proceedings of the 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Cracow: Foundation of the Cracow University of Economics, 183-192.

[16] Kuc, M., (2017a). Is the regional divergence a price for the international convergence? The case of Visegrad group. *Journal of Competitiveness*, 9(4), 50-65.

[17] Łyszczarz, B. (2016). Public-private mix and performance of health care systems in CEE and CIS countries. *Oeconomia Copernicana*, *7*(2), 169-185.

[18] Mandić, K., Delibašić, B., Knežević, S. & Benković, S. (2017). Analysis of the efficiency of insurance companies in Serbia using the fuzzy AHP and TOPSIS methods. *Economic Research-Ekonomska Istraživanja*, 30(1), 550-565.

[19] Mardani, A., Jusoh, A., Zavadskas, E.K., Kazemilari, M., Ungku, N.U.A., & Khalifah, Z. (2016). Application of Multiple Criteria Decision Making Techniques in Tourism and Hospitality Industry: a Systematic Review. *Transformations in Business & Economics*, 15(1)(37), 192-213.

[20] Meluzín, T., Balcerzak, A.P., Pietrzak, M. B., Zinecker, M., & Doubravský, K. (2018). The impact of rumours related to political and macroeconomic uncertainty on IPO success: evidence from a qualitative model. *Transformations in Business & Economics*, 17, 2(44), 148-169.

[21] Meluzín, T., Pietrzak, M. B., Balcerzak, A. P., Zinecker, M., Doubravský, K., & Dohnal, M. (2017). Rumours Related to Political Instability and their Impact on IPOs: The Use of Qualitative Modeling with Incomplete Knowledge. *Polish Journal of Management Studies*, 16(2), 171-187.

[22] Meluzín, T., Zinecker, M., Balcerzak, A.P., Doubravský, K., Pietrzak, M. B., & Dohnal, M. (2018). The timing of initial public offerings – non-numerical model based on qualitative trends. *Journal of Business Economics and Management*, 19(1), 63-79.

[23] Müller-Frączek, I. &, Pietrzak, M. B. (2011). Spatial Analysis of the Unemployment Rate in Poland in 2004-2008. In P. Jedlicka (Ed.). *Hradec Economic Days 2011*. Hradec Králové: Gaudeamus, 205-209.

[24] Nyga-Łukaszewska, H., & Chilimoniuk-Przeździecka, E. (2017). Modelling Energy Security and International Competitiveness: The Export Perspective. *Entrepreneurial Business and Economics Review*, 5(2), 71-84.

[25] Pieloch-Babiarz, A. (2017). Determinants of payout policy and investment attractiveness of companies listed on the Warsaw Stock Exchange. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(4).

[26] Pietrzak, M. B. (2016). The Problem of the Inclusion of Spatial Dependence Within the TOPSIS Method. *Montenegrin Journal of Economics*, 12(3), 69-86.

[27] Pietrzak, M. B. (2017). Structural Equation Modeling of Regional Economic Development in Polish Voivodeships in the Years 2010-2014. In T. Loster & T. Pavelka (Eds.). *The 11th International Days of Statistics and Economics. Conference Proceedings. September 8-10, 2016*. Prague: Libuse Macakova, Melandrium, 1214-1222.

[28] Pietrzak, M. B., Balcerzak, A. P., Gajdos, A., & Arendt, Ł (2017a). Entrepreneurial environment at regional level: the case of Polish path towards sustainable socio-economic development. *Entrepreneurship and Sustainability Issues*, 5(2). 190-203.

[29] Pietrzak, M. B, Gajdos, A., & Arendt Ł. (2017b). Assessment of Socioeconomic Development in Poland in the years 2010-2015. In T. Kliestik (Ed.). *17th 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, 4th – 5th October 2017. (Part IV.).* Zilina: University of Zilina, 1951-1956.

[30] Pietrzak M. B, Wilk J., & Siekaniec M. (2013). The Impact of Metropolitan Areas on Internal Migrations in Poland. The Case of Southern Regions. In Papież, M., Śmiech, S. (eds.), Proceedings of the 7th Professor AleksanderZelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Cracow: Foundation of the Cracow University of Economics, 124-132.

[31] Pietrzak M.B., Wilk J., Kossowski T. & Bivand R. (2014). The Identification of Spatial Dependence in the Analysis of Regional Economic Development - Join-count Test Application. In M. Papież & S. Śmiech (Eds.). *Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena.* Cracow: Foundation of the Cracow University of Economics, 135-144.

[32] Pietrzak, M. B, & Ziemkiewicz, B. (2017). The use of random fields in the Modifiable Areal Unit Problem. In M. Papież and S. Śmiech (Eds.). *The 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings.* Cracow: Foundation of the Cracow University of Economics, 319-328.

[33] Rogalska, E. (2017). Multivariate analysis of entrepreneurship in Poland at regional level. In T. Kliestik (Ed.). *17th 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, 4th – 5th October 2017. (Part V.).* Zilina: University of Zilina, 2198-2202.

[34] Stanickova, M. (2017). Can the implementation of the Europe 2020 Strategy goals be efficient? The challenge for achieving social equality in the European Union. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(3), 383–398.

[35] Shpolianska, I., Dolzhenko, A., & Glushenko, S. (2017). Fuzzy models of web services selection in the development of service-oriented information systems for small and medium enterprises, *Ekonomicko-manazerske Spectrum*, 11(2), 64-75.

[36] Szyja, P. (2016). The role of the state in creating green economy. *Oeconomia Copernicana*, 7(2), 207-222.

[37] Stankeviciene, J., Kraujaliene, L. & Vaiciukeviciute, A. (2017). Assessment of technology transfer office performance for value creation in higher education institutions. *Journal of Business Economics and Management*, *18*(6), 1063-1081.

[38] Vavrek R., Adamisin P., & Kotulic R. (2017). Multi-criteria evaluation of municipalities in Slovakia – case study in selected districts. *Polish Journal of Management Studies*, 16(2), 290-301.

[39] Vilys, M., Jakubavičius, A., & Žemaitis, E. (2015). Public Innovation Support Index for Impact Assessment in the European Economic Area. *Entrepreneurial Business and Economics Review*, 3(4), 123-138.

[40] Walczak D. & Pietrzak, M. B. (2016). Analysis of Agrarian Structure in Poland in 1921 and 2002 based on the Example of Selected Districts. In M. H. Bilgin, H. Danis, E. Demir & U. Can (Eds.). *Business Challenges in the Changing Economic Landscape - Vol. 1. Proceedings of the 14th Eurasia Business and Economics Society*. Springer International Publishing, 461-472.

[41] Walesiak, M. (1999). Distance Measure for Ordinal Data. *Argumenta Oeconomica*, 2(8), 167-173.

[42] Walesiak, M., & Dudek A. (2016). Package 'clusterSim', Searching for Optimal Clustering Procedure for a Data Set, R package version 0.45-1. Retrieved form http:// cran.r-project.org/.

[43] Zygmunt, A. (2017). Innovation activities of Polish firms. Multivariate analysis of the Moderate Innovator countries. *Oeconomia Copernicana*, 8(4), 505–522.

[44] Żelazny, R., & Pietrucha, J. (2017). Measuring innovation and institution: the creative economy index. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(1), 43-62.

Resource Allocation and Motivation Strategies in Project Management

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

Abstract. While scheduling, rescheduling or levelling resources in project management plenty of methods were derived in the past. Our approach consists of multiple criteria approach to resource assigning and rewarding based on motivation, monitoring and on resource criticality followed by three basic and one combined decision-making strategies. First, in the article, a model of multiple criteria analysis of variants is derived according to the procedure commonly known in project management praxis, then four new approaches (strategies) are suggested for the project manager. Each of them focuses different parameter: 1. frequency of resource criticality - direct impact on success and strong resource motivation, 2. absolute frequency of resource allocations - hidden impact on success, prevention of risk, and 3. absolute amount of resource work in a project – emphasis on resource cooperation and group motivation. A specific mathematical model supports each of these strategies. Our article links the field of project management and multi-criteria decision making in the context of current practice and proposes its own process and mathematical model for solving the problem.

Keywords: resource management, resource criticality, motivation strategy, utility function.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Project managers face a series of difficulties in the process of managing projects. Importantly, these include: a wide variety of activities performed by people of different abilities during the life cycle of the project; the complexity of the management process, which evolves a series of interrelated activities, the uncertainties surrounding the process of project management, and the external, dynamic environment [4] with frequent failures [1] or [2]. One of these difficulties is definitely the right motivation of working resources to achieve project goals, always taking into account the individual characteristics of individual workers, the importance of their work for the project in the context of other activities and other resources.

For this purpose, it is possible to use both pure quantitative tools, based on given source, task, project, and individual soft, qualitative tools. This process needs to carry out (control, manage) several tasks simultaneously, performed by several people from different areas of expertise and there is often inadequate knowledge of the tasks to be performed [4]. Scheduling and leveling resources, using different types of management strategies, continues to develop the theory Complex Project Management [6] or [7]. The authors' intention is to derive a mathematical model for the utility function of individual motivational strategies and to measure the impact on project objectives, knowing that the results of the model will have always to be applied ad hoc within the decision making process.

2 Materials and methods

2.1 Project Schedule Management

The project preparation and planning process is described in particular by international standards, e.g. [5] and Kerzner [3]. The initialization phase of the project schedule always begins with the identification of activities and a list of resources. The relationship between project activities and allocated resources is incidental - each resource is allocated at a certain amount of work for the activity, and each activity has at least one allocated resource with the number of units assigned to it. Estimates of the duration of the activity are then given. In the case of parallel assignment following formula sets task duration:

[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.

$$t_i = max\left(\frac{v_i^k}{r_i^k}\right) \tag{1}$$

where t_i is duration of task i, v_i^k is the amount of resource k work allocated to task i, r_i^k is the number of units of resource k assigned to task i.

In case of a non-parallel allocation, the project manager has the possibility to apply serial resource assignment, to use overlays, different work contours, task splitting, etc. [3] according to the current resource parameters or individual needs. In the case of serial assignment following formula sets task duration:

$$t_i = \sum \frac{v_i^k}{r_i^k} \tag{2}$$

Combination of both assignment types is off course possible as so as using of overlays (lead times) or lags between subsequent assignments. Moreover parameter r may not be a constant and it can be defined as a function of task duration (decreasing or increasing or simply changing in time). Following formula describes situation, where resource assignments is constant only within a given period of time m^k (e.g. work contour of "back loaded", "front loaded" or similar types).

$$t_i = \max_k \left(\frac{m^k v_i^k}{\sum_j^{m^k} r_{ij}^k} \right) \tag{3}$$

The sequence of activities in the project plan is given not only by technological links, but also by resource feasibility [5]. The timetable for the project is always drawn up in such a way that the project is carried out as soon as possible, with all the linkages and limitations, if possible with respect to the allocated budget. This fact needs to be accepted while proper resource scheduling.

2.2 Project Human Resource Management

Project Human Resource Management consists of organizing and managing project and resource team, or managing other stakeholders involved in the project [5]. The key steps in this process are: planning, obtaining, allocating and developing resources. There is always an emphasis on motivation and control of resources. The aim is to ensure accountability at all levels of the project, with a direct link to the work tasks and planned outputs of the project.

The most important aspect for obtaining and sustaining resources in a project is motivation. Motivation of resources should be proportional to their involvement in the project - it is necessary to stimulate both the individuals and the team positively. Although financial and non-financial motivation can be distinguished, financial instruments are often preferred in projects. The method of redistribution of rewards is based on the decision of the project manager. Motivation of resources using financial rewards can be differentiated according to the approach, either with an emphasis on individuals or an emphasis on the entire team [3].

3 Results and Discussion

3.1 Motivation Strategies in Project Management

Proper motivation of resources in a project makes achieving project goals more likely, both in quantity and quality. Every resource within the project or the project team requires other motivational tools. Thus, the project manager has to reconcile the personality traits of the individual with his / her role in the project team, i.e. his / her importance for successful project completion. Such importance can be expressed and measured in different ways and it is possible to define several motivational strategies (MS) according to different indicators, in principal based on two so called Control strategies (CS). All strategies are derived from the multi-criteria utility function, where this function always includes all criteria but with different preferences:

Control strategy 1 - "**Direct Impact on Success**". This strategy emphasizes criticality of the resource (possible threat to deadlines) with the inclusion of allocation and total labor. As mentioned above, resource allocation to a task can be either serial (and thus any resource used is critical and crucial for in time completion of the task) or parallel, where the resource is critical, if it determines task duration and so it is crucial for its completion, i.e. $t_i = \frac{v_i^k}{r_i^k}$. "Direct Impact on Success" assumes further motivation for timely termination of a task without a direct link to other ongoing tasks. The criterion for this strategy is therefore resource criticality and frequency of such criticality.

Control strategy 2 - "**Hidden Impact on Success**". This strategy emphasizes the frequency of resource allocation throughout the project regardless criticality:

$$f_k = \frac{n_k}{N} \tag{4}$$

where n_k is the number of tasks where the resource is allocated and N is total number of project tasks. Next criterion for this strategy is the total amount of work for the resource throughout the project i. e. :

$$\sum_{i=1}^{N} v_i^k. \tag{5}$$

This strategy reflects the danger of the crisis resulting from aggregation of risks or critical masses - for example, a sudden outage of a resource that was not critical at all but provided work on most tasks. Criteria for this strategy are thus two - the frequency of allocation and the amount of resource work.

The purpose of the control strategies is to determine the frequency and importance of controls, or the degree of intensity or depth of them. According to the results, the source is further motivated - rewarded.

Based on control strategies, managerial motivation can be applied. Their use depends on the type of project and the situation in which it is realized.

Motivational strategy 1 - "**Winner Takes Everything**". This strategy influences the style of resource management. It puts a strong emphasis on the criterion of criticality of the resource (the crisp motivation of the individuals). The success of the project as a whole is purely derived from a cumulative success of all individual tasks.

Motivational strategy 2 - "**Everyone's merit of the whole**". This is again a strategy that influences the style of resource management. It cumulates all three criteria (criticality, amount of work, frequency of assignment) and proportionally distributes rewards to resources. Strategy is good for motivating the team as a whole.

3.2 Model of Utility for Resource Allocation

As described above, resource k will be assessed according the following three criteria:

- number of tasks where it is critical (let it be denoted by c_k);
- number of tasks where it is allocated n_k;
- total amount of work for the resource throughout the project (let it be denoted by $a_k = \sum_{i=1}^{N} v_i^k$).

Next, let us introduce the following notation:

$c_H = \max_k c_k$, $n_H = \max_k n_k$, $a_H = \max_k a_k$, $c_D = \min_k c_k$, $n_D = \min_k n_k$, $a_D = \min_k a_k$.

Now we can define normalized values of criteria for resource k

$$\bar{c}_k = \frac{(c_k - c_D)}{3(c_H - c_D)}, \bar{n}_k = \frac{(n_k - n_D)}{3(n_H - n_D)}, \bar{a}_k = \frac{(a_k - a_D)}{3(a_H - a_D)} \tag{6}$$

so that expression $\bar{c}_k + \bar{n}_k + \bar{a}_k$ represents a linear utility function with the same weights 1/3 (thus 3 in denominator in the normalization formulas) for all criteria acquiring values between 0 and 1. In the reality weights can vary and they can be derived based on survey among project managers but obtaining relevant real data about weights is out of scope of this paper and finally it will not influence principles of proposed method. If there existed a resource with the best values of all criteria within the whole project or a resource with the worst values of all criteria within the whole project, the value of this utility function would be 1 or 0, respectively.

When constructing utility functions for single strategies defined in the previous section, we will use degressive partial (one-criterion) utility functions for criteria with strong emphasis put on and progressive partial utility functions for unimportant criteria. The following function shows to be suitable for this purpose:

$$s_k = \sqrt[\omega_c]{\bar{c}_k} + \sqrt[\omega_n]{\bar{n}_k} + \sqrt[\omega_a]{\bar{a}_k} \tag{7}$$

Parameters ω_c, ω_n, ω_a characterize importance of corresponding criteria. We use not powers but roots to obtain relation the bigger the emphasis on the criterion is, the bigger the parameter is, too. Note that the partial utility function is degressive if the parameter is greater than 1 and progressive in the opposite case. The last step of construction of general formula for the utility function is its normalization so that its values will mean the part of overall attention within controls or of the total amount of finances for rewarding paid to a single resource:

$$u_k = \frac{s_k}{\sum_i s_i} \tag{8}$$

For single motivational strategies described above, we propose the following parameters in the utility function:

- Control strategy 1: $\omega_c = 3$; $\omega_n = 1$; $\omega_a = 1$;
- Control strategy 2: $\omega_c = 0.5$; $\omega_n = 1$; $\omega_a = 0.5$;
- Motivational strategy 1: $\omega_c = 1$; $\omega_n = 1/3$; $\omega_a = 1/3$;
- Motivational strategy 2: $\omega_c = 1$; $\omega_n = 1.25$; $\omega_a = 2.5$.

3.3 Case Study:

For the purposes of this paper and authors' proposals given above the following illustrative case study of small extension project accompanied with resource allocation was chosen. Each task is ensured by several resources, each resource having a number of units equal to 1:

duration	*8 days*	*7 days*	*5 days*	*8 days*	*5 days*	*8 days*	*9 days*	*8 days*
	Task 1	**Task 2**	**Task 3**	**Task 4**	**Task 5**	**Task 6**	**Task 7**	**Task 8**
Thomas	8		5					8
John	5		4			6		6
Peter	4		2	6		5		2
Barbara		5		8		8		
George		6			5		9	
Anna	7				4		7	6
Theresa		7		6				4

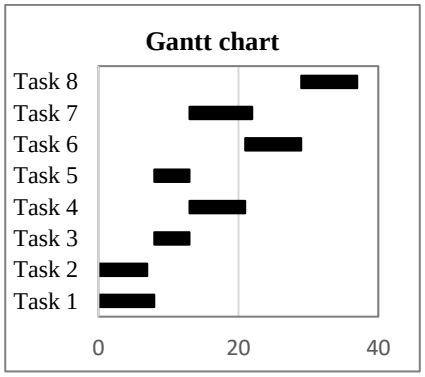


Table 1 Case study: Resource allocation and task duration with Gantt Chart

The estimated time of task duration is presented in the top row of the table. The resource who determines the time of the task duration by his/her amount of work (maximum value in the column) is critical for the given task – its outage or delay would endanger the task duration time. In contrast, this fact is not usually emphasized in the Gantt Chart.

For the presented case study, it is possible to derive the following multiple-criteria decision-making model. Its criteria capture varied nature of the resource allocation in the project: a resource who has low total amount of work may be critical (most often determine the task duration time) and vice versa:

	Frequency of resource criticality	Frequency of allocation	Total amount of work in Man-days
Thomas	3	3	21
John	0	4	21
Peter	0	5	19
Barbara	2	3	23
George	2	3	20
Anna	0	4	24
Theresa	1	3	17

Table 2 Multiple-criteria model: Resource criticality and allocation

When coordinating the project, the task for the project manager is to ensure sufficient resource motivation. Three criteria derived and presented here make this task to be nontrivial. Therefore, the next step is to use the

proposed mathematical models as supporting tool for project manager decisions. Table 3 shows normalized criteria, Table 4 partial benefits for individual resources according to different strategies.

	Frequency of resource criticality	Frequency of allocation	Total amount of work
Thomas	0,333333	0	0,190476
John	0	0,166667	0,190476
Peter	0	0,333333	0,095238
Barbara	0,222222	0	0,190476
George	0,222222	0	0,142857
Anna	0	0,166667	0,333333
Theresa	0,111111	0	0

Table 3 Normalized criteria table

	CS1	**CS2**	**MS1**	**MS2**
Thomas	0,21	0,13	0,34	0,18
John	0,09	0,18	0,01	0,16
Peter	0,10	0,30	0,04	0,17
Barbara	0,19	0,08	0,23	0,15
George	0,18	0,06	0,23	0,14
Anna	0,12	0,24	0,04	0,18
Theresa	0,11	0,01	0,11	0,02

Table 4 Resources utility values computed for single strategies

Resulting utility for individual resources can be used as a weight for dividing the reward or for determining the number or frequency of controls in the project. The selection of utility - benefits according to the strategy (its type) leads to facilitating the decision of the project manager - the choice of strategy can be made based on the nature of the project, the resources of the team or the situation during the project implementation. The project manager, through alternating strategies, can either support the team as a whole or highlight the results of the individual resource as a member of a team, and thus flexibly respond to the changing situation in a project.

4 Conclusion

Four basic control and motivational strategies for more efficient use of resources in the project were presented in the article, aiming to bring the project objectives and goals as close as possible. A strategy of focusing on a critical (key) individual resource and its role in fulfilling a unique task does not play a major role, as was expected in high frequency criticality sources, however, their motivation is crucial for the project (value of utility function for MS1). A strategy of focusing on team motivation (MS2) is thus important for highly allocated resource.

Further research: Motivational strategies based on three core criteria will be verified in corporate practice. The authors furthermore predict the specification of the criterion of criticality of a resource. In practice, it is relevant that resources may not be assigned in parallel - they may be assigned in serial (where one resource takes over the work of its predecessor and all resources are now critical) or assigned with overlay where the resource is critical only for a particular phase of the task. The criticality criterion will then need to be modified but not in the whole. Practicing serial allocation divides a task into subtask each with one critical resource in above described meaning. Further development is expected in terms of resource work and critical mass. Work can be relativized due to the workload or normalized to the work of the whole project. Critical mass then threatens resources assigned to the routine or turnaround, in the meantime, of minor activities. Accumulation of their failure may jeopardize the success of the whole project.

References

[1] Cooke-Davies, T. (2002). The “real” success factors on projects. *International Journal of Project Management*, *20*, 185–190.
[2] Hameri, A. P. (1997). Project management in a long-term and global one-of-a-kind project. *International Journal of Project Management*, *15*, 151–157.
[3] Kerzner, H. (2013). *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*. Eleventh edition. New York: John Wiley & Sons Inc., p. 1264. ISBN 978-1-118-02227-6.
[4] Mota, C. M. M., de Almeida, A.T. (2011). A multicriteria decision model for assigning priority classes to activities in project management. *Annals of Operations Research*, *199*, 361–37.
[5] Project Management Institute (2017). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*. Sixth Edition. Newtown Square: Project Management Institute Inc. ISBN 978-1-62825-184-5.
[6] Whitty, S. J., Maylor, H. (2009). And then came Complex Project Management (revised). *International Journal of Project Management*, *27*, 304–310.
[7] Williams T. (1999). The need for new paradigms for complex projects. *International Journal of Project Management*, 17, 269–73.

Differences in income distribution given by the structure of household

Jitka Bartošová[1], Vladislav Bína[2]

Abstract. Incomes of households reflect the socioeconomic development of population and simultaneously influence the state of society because the incomes affect the behavior of households as the basic socioeconomic subjects. A direct relation exists between incomes and purchasing power which to a significant extent determines the consumption behavior of households.
The presented paper is focused on specification of differences in the income distribution according to their structure. Since the different types of households mutually differ not only in their size and number economically active members but also in their demographic structure (age, education, etc.) we can expect differences also in the character of their income distributions. The goal of the paper is a construction of appropriate income distribution models and identification of differences. For this purpose, we can use both basic sample characteristics and suitable models of frequency distribution. The modeling of frequency distribution can be performed using non-parametric methods (kernel estimates with Gaussian kernel) and parametric models (finite mixtures of densities). The finite mixtures of densities are usually considered as more suitable for modeling of distributions of random variables in case of inhomogeneous populations than simple lognormal models with two or three parameters.

Keywords: income distribution, finite mixtures of densities, household structure.

JEL Classification: C33, C44
AMS Classification: 62E17

1 Models of empirical frequency distribution

Role of a model generally lies in the simplification and generalization of reality. Usually, we distinguish between parametric and non-parametric models. The non-parametric description of distribution is based on empirical values and does not need any a priori information concerning unknown (modeled) quantity. The result is a non-parametric curve which (in dependence on smoothing parameter) in a more or less detailed way express the information contained in observed (empirical) values. The parametric description of estimated quantity (in our case income distribution) we use parametric curves or their mixtures. The parametric models provide higher level of generalization of reality and facilitate to perform a reconstruction of situation for different combination of values of observed variables and predict the future evolution.

1.1 Parametric and non-parametric density models

As non-parametric density estimates of empirical distribution of frequencies the histograms and kernel estimates are used. The kernel estimate of density $\hat{f}_n(x),\ x \in R$, is given by the formula

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n} \right), \tag{1}$$

where K is a kernel of the estimate and h_n is so called smoothing parameter. Its choice has an impact on the shape of kernel and degree of concentration in point X_i on the neighborhood. The most frequently used kernel applicable also on the case of empirical distribution of incomes is the normal (Gaussian) kernel with the formula

$$K\left(\frac{x - X_i}{h_n} \right) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{x - X_i}{h_n} \right)^2 \right]. \tag{2}$$

[1,2] University of Economics in Prague, Faculty of Management, Department of Exact Methods, Jarošovská 1117/II, 37701 Jindřichův Hradec, {bartosov,bina}@fm.vse.cz.

The basis of parametric modeling is a normal distribution with its symmetric (Gaussian) density function. But the empirical income distribution is rather asymmetric and corresponds rather to the lognormal density. For this reason, it is advantageous to employ logarithms of incomes and to approximate it by normal distribution. Under the assumption that the incomes can be approximated by two-parametric lognormal distribution with parameters μ and σ^2, their logarithms are distributed according to the normal distribution $N(\mu, \sigma^2)$ where μ and σ^2 are mean and variance on the logarithmic scale.

But the empirical distribution is quite frequently considerably heterogeneous and thus, the simple parametric models are not able to approximate the distribution with sufficient accuracy. In such case it is more advantageous to employ finite mixtures of densities.

The model of finite mixture with K components we generally assume that the observed population P represented by the date from the analyzed sample compose of K particular subpopulations (subgroups, or subsets) P_k, $k = 1, \dots, K$ and each of them is described by a distribution (or its model). By mixture (parametric) model we understand a set of distributions with one or more unknown parameters which need to be estimated. Density of a multidimensional random variable X is thus a finite mixture of K densities if

$$f^{(K)}\left(X; \mathbf{\Psi}^{(K)}\right) = \sum_{k=1}^{K} p_k^{(K)} f_k^{(K)}\left(\mathbf{x}; \mathbf{\Theta}_k^{(K)}\right), \tag{3}$$

where $p_k^{(K)} > 0$ for $k = 1, \dots, K$ and $\sum_{k=1}^{K} p_k^{(K)} = 1$ are weights (marginal probabilities) of components, $f_k^{(K)}\left(\mathbf{x}; \boldsymbol{\theta}_k^{(K)}\right)$ are densities of particular components and $\mathbf{\Psi}^{(K)} = (p_1^{(K)}, \dots, p_{K-1}^{(K)}, \boldsymbol{\theta}_1^{(K)}, \dots, \boldsymbol{\theta}_K^{(K)})$ is a vector of unknown mixture parameters (in detail see, e.g., McLachlan and Peel in [11]).

For the estimation of mixture parameters EM-algorithm is used [6] which uses the well-known fact that every continuous distribution can be approximated with sufficient accuracy by the mixture of densities of Gaussian distributions. EM-algorithm is based on a principle of Bayesian clustering. Bayesian clustering is one of non-hierarchical (optimization) clustering methods requiring to set the number of cluster (mixture component K) to be given in advance. But since the number of components is not known before, it is necessary to infer it using some penalized information criterion – usually Akaike information criterion AIC [1] or Bayesian information criterion BIC [13] is used. In agreement with the parsimony principle, the notion of penalization means a "penalty" for the increasing of the number of model parameters, i.e. the growth of its complexity. Criteria do not bear information about the ability of model to depict the reality and cannot be used to test the statistical significance of the model. They can be used just for the comparison of pair of models and to find out the difference of information loss under the assumption that we choose some model instead of some other.

EM-algorithm is implemented in the *mclust* package (see [14]) which is one of the packages of R software [12] and is used for the estimation of parameters of the mixture and able to employ SILC observation weights correctly.

2 Characterization and modeling of empirical distribution of Czech households' income

Estimation of basic characteristics of income distribution and construction of suitable models of incomes and wages in Czech households (or individuals) is studied in many papers. Characterization of distribution properties, its development and future prediction is a focus of papers, e.g., by Bílková [5] or Marek [9]. Other papers (like [4]) discuss possibilities of estimation of empirical distribution using simple lognormal model with two or three parameters. This simple parametric estimate can be augmented by a more sophisticated and precise type – an estimation using finite mixtures of normal distributions (see, e.g., papers [2], [3], [7], [8], [10]. These papers are devoted to the models of income or wage distribution as an aggregate without consideration of structure of particular households. Our goal in the presented paper is to choose more detailed approach and construct suitable models on the level of particular structural type of households. The division of households into the groups is performed using typology defined in EU (see Table 1).

Household type	Structure	Household type	Structure
Type 0	individuals	Type 5	2 adults, 1 child
Type 1	2 adults, both under 65	Type 6	2 adults, 2 children
Type 2	2 adults, at least one 65+	Type 7	2 adults, 3 and more children
Type 3	other h. without children	Type 8	other households with children
Type 4	1 adult, 1 or more children	Type 9	other (not typical) households

Table 1 Structure of households according to typology defined and used in EU.

The income distribution can be seen from three types of perspectives. From the perspective of household as a basic unit, from the perspective of income per individual, i.e. total income divided by number of household members, or from the perspective of so called equalized income, i.e. total income divided by number of consumption units[2]. In our paper we will employ the first type of perspective and we will consider total incomes of households where higher heterogeneity could be assumed than in case of incomes per individual or per consumption unit.

The data base comprises data from Czech variant of sample survey of EU-SILC "Living conditions" from 2015. It is a large random sample survey performed each year in all EU member states and its results can be rather straightforwardly generalized.

2.1 Basic characteristics of logarithmic income for different household types

For the description of location and variability of incomes in logarithmic scale we used basic robust and classical characteristics – namely mean, median, standard deviation and interquartile range. Results for all households and particular household types in 2015 are summarized in Table 2. Types of households are sorted in a decreasing way according to the value of logarithmic mean income of households.

Household type	Mean	Standard deviation	Median	Interquartile range	Number of households
All households	**12,66**	**0,61**	**12,68**	**0,84**	**7914**
Type 0	13,16	0,41	13,15	0,48	423
Type 5	13,16	0,40	13,15	0,48	1018
Type 8	12,98	0,60	13,01	0,86	99
Type 7	13,01	0,46	13,01	0,47	510
Type 6	12,95	0,49	12,95	0,48	428
Type 3	12,80	0,46	12,81	0,57	1299
Type 4	12,61	0,29	12,56	0,28	1441
Type 9	12,31	0,47	12,33	0,58	197
Type 1	12,15	0,55	12,16	0,68	1105
Type 2	11,94	0,25	11,90	0,25	1394

Table 2 Basic characteristics of income logarithm for all households and different household types.

It is apparent that maximal incomes – measured either by mean or median – can be observed in households of types 0 and 5, i.e. in households of individuals and 2 adults with one child. It is an interesting, but more or less expectable result since these are households of economically active people. On the other hand, the maximal variability measured by standard deviation and interquartile range in case of type 8, i.e. other households with children is also an expectable result because the number of members and their structure vary significantly in this group. It could be expected that the lowest characteristics of location and variability are observed in case of households of type 2, i.e. households with two adult members where at least one is older than 65 years. According to the fact that these are mostly households of retired people we cannot expect in this group high and varying incomes.

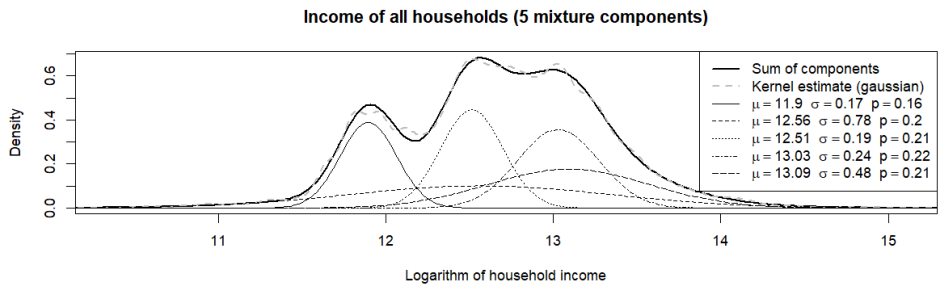


Figure 1 Mixture for all households.

[2] Number of consumption units is lower than the number of household members since it takes into consideration economies of scale, i.e. the savings thanks to the cohabitation in one household.

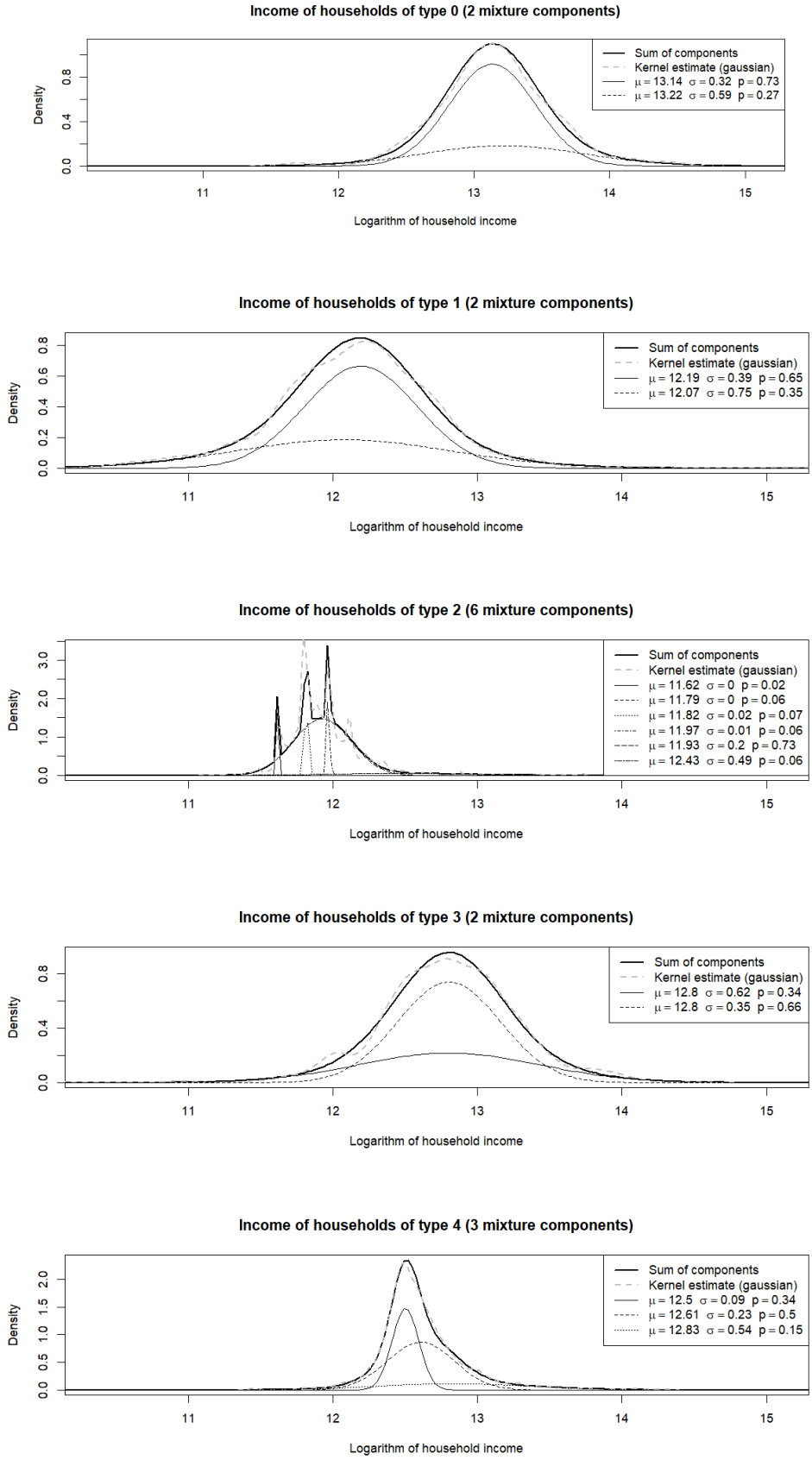


Figure 2 Mixtures for household types 0 to 4.

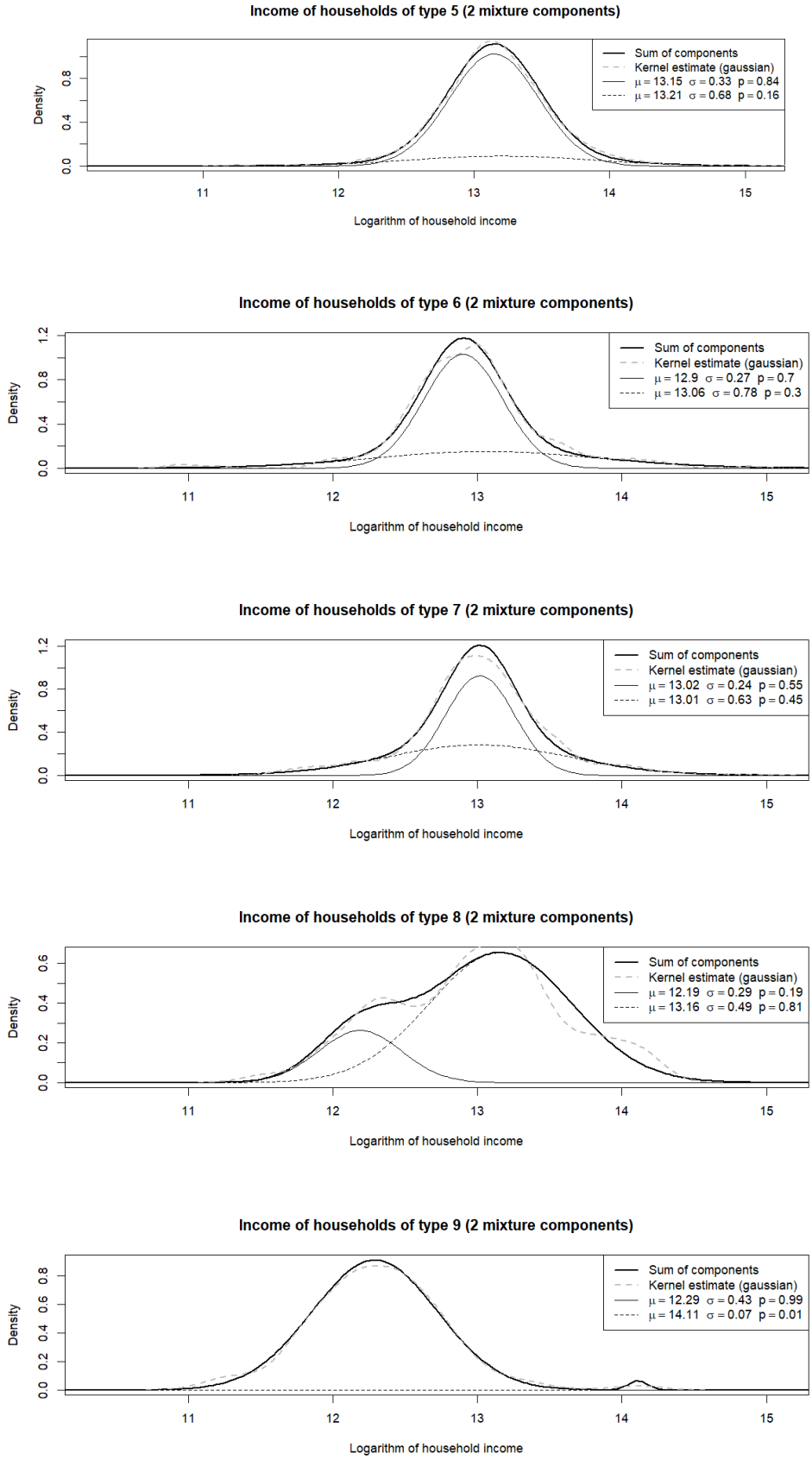


Figure 3 Mixtures for household types 5 to 9.

2.2 Modeling of density by mixture models and comparison with kernel estimates

For approximation of empirical distribution of total net incomes of Czech households, we constructed kernel estimates and finite mixture models of densities with optimal number of components. The results are depicted in Figures 1 – 3. Based on the Akaike information criterion which was used for the optimization of number of components in the mixtures we constructed mixtures with two components (types 0, 1, 3, 5, 6, 7, 8 and 9), three components (type 4), four components (all Czech households regardless of the type) and with six components (type 2). All households together are modeled using five mixture components.

3 Conclusion

The assumption of a considerable heterogeneity of income distribution was fulfilled in case of all households regardless its type and in case of types 2 and 4, i.e. the households with two adult members where one is older than 65 and the households with one adult member and one or more children. These distributions show considerable multimodal character as we can observe from the kernel estimate of density. These are the household types where small groups exist which differ considerably from the income of the majority. In case of other household types (types 0, 1, 3, 5, 6, 7, 8 and 9) we can see that two component mixtures fit the empirical distribution with a sufficient precision. In this case the empirical distribution cannot be considered as heterogeneous, the other way around it is rather homogeneous. It is apparent that the heterogeneity of the empirical distribution is related neither to the location of distribution nor to the variability of the data file.

Acknowledgements

Supported by Institutional support of the long-term development (Faculty of Management in Jindřichův Hradec, University of Economics, Prague) under grant number IP 600040.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov & Csaki (Eds.), *Proceedings International Symposium in Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.

[2] Bartošová, J., & Bína, V. (2014). The measure of separation of components in finite mixtures of probability densities. In J. Talašová, J. Stoklasa & T. Talášek (Eds.) *32nd International Conference Mathematical Methods in Economics, MME 2014, Conference Proceedings* (pp. 13–18). Olomouc: Palacký University.

[3] Bartošová, J. & Longford, N.T. (2014). A study of income stability in the Czech Republic by finite mixtures. *Prague Economic Papers, 23*, 330–348.

[4] Bílková, D. (2011). Analysis of the Development in Wage Distributions of Men and Women in the Czech Republic in Recent Years. *Statistika – Journal for Economy and Statistics, 48*, 40–57.

[5] Bílková, D. (2012). Recent Development of the Wage and Income Distribution in the Czech Republic. *Prague Economic Papers 21*, 233–250.

[6] Dempster, A., Laird, N. & Rubin, D. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39,** 1, 1–38.

[7] Malá, I. (2013). Použití konečných směsí logaritmicko-normálních rozdělení pro modelování příjmů českých domácností. *Politická ekonomie, 61*, 356–372.

[8] Malá, I. (2014). A multivariate mixture model for incomes of the Czech households in 2006–2010. In T. Löster & T. Pavelka (Eds.) *8th International Days of Statistics and Economics* (pp. 656-664). Slaný: Melandrium.

[9] Marek, L. (2017). The effect of extreme wages on average wage values. In T. Löster & T. Pavelka (Eds.) *11th International Days of Statistics and Economics* (pp. 962–969). Slaný: Melandrium.

[10] Marek, L. & Vrabec, M. (2013). Probability models for wage distributions. In H. Vojáčková (Ed.) *31st International Conference on Mathematical Methods in Economics* (pp. 575-581). Jihlava: College of Polytechnics.

[11] McLachlan, G. & Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.

[12] R Core Team (2017). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[13] Schwarz, G. E. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 2, 461–464.

[14] Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models The R Journal 8/1, 205–233.

Histogram Bin Width Selection Methods Implemented in Kernel Density Smoothing Parameter Selection

Aleksandra Baszczyńska[1]

Abstract. Histogram and kernel density estimator are the two most known and frequently used nonparametric estimation methods. In construction of these estimators there is a need to choose the optimal value of bin width (in the case of histogram) or the optimal value of the smoothing parameter (in the case of kernel density estimation).
In paper the properties of classical frequency histogram are presented with the special issue of methods for choosing the bin width optimal value. Sturges', Scott's and Freedman-Diaconis' rules and their modifications are discussed.
The results of simulation study on comparing two nonparametric estimation procedures are presented with the special remarks for users on choosing the optimal value of bin width in histogram and smoothing parameter in kernel density estimation. The comparison between histogram bin width selection methods and kernel smoothing parameter selection methods is made.

Keywords: histogram, kernel density estimation, bin width, smoothing parameter.

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

1 Introduction

Histogram and kernel density estimator are the two most known and frequently used nonparametric estimation methods that provide information about behavior of a random variable. The popularity and usefulness of these estimation methods are strictly connected with both convenient interpretation of regarded estimators and their simply construction. In the process of the estimators' construction the researcher has to take decision about the optimal value of bin width (in the case of histogram) or the optimal value of the smoothing parameter (in the case of kernel density estimation). The best approach may mean automatic and simply selection method of regarded parameters that would help the researcher in the construction of histogram or kernel density estimator ([4], [15]) .

The main aim of the paper is to present simply and easy to use selection methods of choosing the optimal value of bin width and the optimal value of smoothing parameter in kernel density estimator. Indicating the difficulties appearing in the applications of these methods may lead to the new method of choosing the optimal value of parameters. The selection methods are compared by using simulation methods and practical hints for the users are presented, based on the results of the simulation study.

In the further considerations the following assumptions are made ([3], [[6], [7], [13]:

- the density $f(x)$ of the random variable X is unknown;
- a set of n observations $\{X_i\}_{i=1}^n$ is treated as the realizations of independent, identically distributed random variables with the unknown density f;
- the aim is to estimate in a nonparametric way the density on the basis of the observations $\{X_i\}_{i=1}^n$ using histogram and kernel density estimator;
- histogram is defined as follows ([2], [8], [9]):

$$ {}_H\hat{f}_{h_H}(x) = \frac{1}{nh_H}\sum_{i=1}^{n}\sum_j I(X_i \in B_j)I(x \in B_j), \tag{1}$$

where B_j are half-opened intervals called bins $B_j := [x_0 + (j-1)h_H, x_0 + jh_H)$, x_0 is the origin of the histogram (the origin of the histogram is set to be equal 0 for bounded random variables – data values in the sample are nonnegative or equal the sample smallest observed value $x_0 = a$ for the data values fall into the support interval (a, b)), h_H $(h_H > 0)$ is the histogram bin width, $j = -\infty, \dots, -1, 0, 1, \dots + \infty$;

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

- kernel density estimator is defined in the following way ([2], [5]):

$$_K\hat{f}_{h_K}(x) = \frac{1}{nh_K}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_K}\right), \quad (2)$$

where K is the kernel function and h_K ($h_K > 0$) is the kernel bindwidth called smoothing parameter. Some examples of kernel functions and the main properties of the kernel density estimator are described in [1], [2], [12] and [14].

- both histogram and kernel density estimator are constructed with continuously measured data.
- histogram is assumed to be equally spaced (with equal-width bins); kernel density estimator is assumed to have fixed smoothing parameter.

The bin width for the histogram and the smoothing parameter for the kernel density estimator play very similar role in constructing the estimators. The values of these parameters imply the degree of smoothing in the estimate. For small values we can get very noisy estimators – it is even a needleplot $h_H \to 0$ or $h_K \to 0$. When big values are used it results in very smooth estimators, even it can be very flat estimator. It means that the problem of selection the proper value of the histogram bin width and the kernel smoothing parameter is of crucial role.

2 Histogram bin width selection methods

The optimal construction of the histogram involves the discussion of two parameters that specify completely the histogram: the origin of the histogram x_0 and the bin width h_H (comp. (1)).

The simplest histogram bin width selection methods are called rules of thumb or normal-reference rules, they are applied when Gaussian basis functions are used to approximate the data. The most often used methods are presented below.

Sturges' bin width rule (based on Sturges' number-of-bins rule) which is a normal-reference rule and it is defined in the following way:

$$h_{H,St} = \frac{x_{(n)}-x_{(1)}}{1+\log_2 n}, \quad (3)$$

where n is the number of observations and $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ denotes ordered values from the sample. This Sturges' bin width rule is widely implemented in software but it causes that estimator is oversmoothed. There is Doane's modification of (3) when the data are not normally distributed and some additionally classes may be required in the histogram. It can be noticed that for large samples both Struges' and Doane's rules can caused that the histograms are oversmoothed.

Scott's rules for the bin width (which are also normal-reference rules), based on mean integrated squared error as the measure of discrepancy, are:

$$h_{H,ScI} = 3.5Sn^{-\frac{1}{3}} \quad (4)$$

and

$$h_{H,ScII} = 3.4908Sn^{-\frac{1}{3}} \quad (5)$$

where S is standard deviation from the sample.

Freedman and Diaconis' rule, which implemented robust estimator IQ instead of standard deviation as it is in Scott's rules, is as follows:

$$h_{H,FD} = 2IQn^{-\frac{1}{3}} \quad (6)$$

where IQ is the sample interquartile range.

For moderate number of observations, sample rules (3), (4), (5) and (6) give similar results ([10]). Scott's rules which belong to normal reference rule family use normal distribution in the formula of asymptotically optimal bin width when point-wise mean squared error of the histogram is regarded. In the formula of optimal bin width $h^* = \left[\frac{6}{n\int_{-\infty}^{\infty} f'(x)^2 dx}\right]^{\frac{1}{3}}$, in the place of the roughness of the unknown density f ($\int_{-\infty}^{\infty} f'(x)^2 dx$), the roughness of the normal density with variance mean σ^2 of the following form $\int_{-\infty}^{\infty} f'(x)^2 dx = \frac{1}{4\sqrt{\pi}\sigma^3}$ is used [11]. In the Freedman and Diaconis' rule the following approximation of standard deviation is used: whereas for standard normal data the difference $\Phi(0.75) - \Phi(0.25)$ is 1.349, where Φ is cumulative distribution function for stand-

ardized normal distribution, the robust estimator of standard deviation is $\frac{IQ}{1.349}$. So, $3.4908 \cdot \frac{IQ}{1.349} = 2.588$. Moreover, Freedman and Diaconis suggested 2.0 instead of 2.588.

3 Simulation study and the results

The simulation study is conducted to investigate the possibility of using values of histogram bin width in selection of kernel smoothing parameter value. Bin width rules: (3)-(6) are normal reference rule and are rules that are very often used in practical usage because of their simplicity and popularity especially in statistical software. It seems to be very natural approach to implement these rules in more sophisticated density estimator as kernel density estimator. All regarded histogram bin width rules are based on data (because of the estimated measure of deviation used in rules) and are based on mean integrated squared error as some rules for selecting the smoothing parameters in kernel density estimation. The motivation of this simulation study is also connected with the issue of the sample sizes, especially with the question how large the sample to be to optimally select the proper value of histogram bin width and kernel smoothing parameter.

In the study the data sets are randomly taken from population of the stable distribution (α stable distribution) as on one hand the generalization of normal distribution and on the other hand the distribution that is applied widely in economics, especially financial economics. Moreover, it is feasible used in many situations where economic variables are considered. Stable distribution $S(\alpha, \beta, \gamma, \delta)$ is indexed by four parameters: α describes the tails of the distribution (it is called first shape parameter, $0 < \alpha \leq 2$, the bigger α the more extreme values are observed); β describes the skewness of the distribution (it is called second shape parameter, $-1 \leq \beta \leq 1$, for $\beta = 0$ the distribution is symmetric, for $\beta > 0$ the distribution is right-skewed, for $\beta < 0$ the distribution is left-skewed; γ (scale parameter; $0 < \gamma < \infty$) and δ (location parameter; $-\infty < \delta < \infty$).
In the study the following population distributions are taken into consideration:

1. $S\left(2, 0, \frac{\sqrt{2}}{2}, 0\right)$, normally distributed population with mean 0 and variance 1 ($N(0,1)$), population P1;
2. $S(2, 0{,}5\sqrt{2}, 0)$, normally distributed population with mean 0 and variance 100 ($N(0,100)$), population P2;
3. $S(1, 0, 1, 0)$, standard Cauchy distributed population with scale parameter 1 and location parameter 0, population P3;
4. $S(1, 0, 2, 0)$, Cauchy distributed population with scale parameter 2 and location parameter 0, population P4;
5. $S(0.5, 1, 0.5, 1)$, Lévy distributed population with scale parameter 0.5 and location parameter 0.5, population P5;
6. $S(0.5, 1, 4, 4.5)$, Lévy distributed population with scale parameter 4 and location parameter 0.5, population P6.

From the above populations the samples are taken with the sizes $n = 30, 50, 100, 300, 1000$. Different sample sizes widen the study to the situations where small, medium and large samples are considered.

For every sample the values of histogram bin width are calculated using (3)-(6). In addition for the same samples values of the kernel smoothing parameters are calculated (with Gaussian kernel function) using the following smoothing parameter methods: iterative ($h_{K,IT}$), direct plug-in ($h_{K,DPI}$), maximal smoothing ($h_{K,MS}$), biased cross-validation ($h_{K,BCV}$), least squared cross-validation ($h_{K,LSCV}$), reference rule ($h_{K,RR}$). The number of repetition is set to 5000.

Calculated values of histogram bin width and kernel smoothing parameter for the same samples are compared. In this way the possibility of using selection method of bin width in kernel density estimation can be analyzed. Tables 1-6 present histogram bin width and kernel smoothing parameter values for chosen sample sizes.

Sample size	Histogram bin width				Kernel smoothing parameter					
	$h_{H,St}$	$h_{H,ScI}$	$h_{H,ScII}$	$h_{H,FD}$	$h_{K,IT}$	$h_{K,DPI}$	$h_{K,MS}$	$h_{K,BCV}$	$h_{K,LSCV}$	$h_{K,RR}$
30	0.55	1.03	1.02	0.86	0.68	0.59	0.47	0.70	0.73	0.43
50	0.56	0.87	0.86	0.57	0.54	0.53	0.42	0,.57	0.56	0.39
100	0.62	0.78	0.78	0.58	0.49	0.57	0.45	0.54	0.28	0.41
300	0.65	0.56	0.56	0.46	0.38	0.51	0.39	0.42	0.40	0.36
1000	0.65	0.34	0.34	0.26	0.26	0.36	0.28	0.28	0.30	0.26

Table 1 Values of smoothing parameter for samples from population P1

Sample size	Histogram bin width				Kernel smoothing parameter					
	$h_{H,St}$	$h_{H,ScI}$	$h_{H,ScII}$	$h_{H,FD}$	$h_{K,IT}$	$h_{K,DPI}$	$h_{K,MS}$	$h_{K,BCV}$	$h_{K,LSCV}$	$h_{K,RR}$
30	7.82	13.03	13.00	9.63	7.87	7.21	6.19	8.93	7.87	5.73
50	6.86	9.85	9.83	8.06	6.26	6.38	5.42	6.30	7.01	5.02
100	5.78	7.70	7.68	5.70	5.05	5.80	4.48	5.67	1.92	4.15
300	6.12	5.05	5.04	3.90	3.42	4.33	3.52	3.46	3.92	3.26
1000	5.84	3,58	3.57	2.80	5.37	10.68	7.70	5.81	4.95	7.13

Table 2 Values of smoothing parameter for samples from population P2

Sample size	Histogram bin width				Kernel smoothing parameter					
	$h_{H,St}$	$h_{H,ScI}$	$h_{H,ScII}$	$h_{H,FD}$	$h_{K,IT}$	$h_{K,DPI}$	$h_{K,MS}$	$h_{K,BCV}$	$h_{K,LSCV}$	$h_{K,RR}$
30	22.92	26.29	26.22	1.73	1.06	1.47	1.16	4.50	3.23	1.50
50	8.23	6.61	6.59	0.86	0.67	0.78	0.65	1.09	1.09	0.60
100	8.52	4.85	4.84	0,72	0.40	0.62	0.59	0.65	0.65	0.52
300	7.06	3.37	3.36	0.50	0.42	0.68	0.53	1.05	1.05	0.45
1000	118,90	12.79	12.76	0.42	0.31	0.61	0.45	0.90	0.90	0.41

Table 3 Values of smoothing parameter for samples from population P3

Sample size	Histogram bin width				Kernel smoothing parameter					
	$h_{H,St}$	$h_{H,ScI}$	$h_{H,ScII}$	$h_{H,FD}$	$h_{K,IT}$	$h_{K,DPI}$	$h_{K,MS}$	$h_{K,BCV}$	$h_{K,LSCV}$	$h_{K,RR}$
30	7.99	9.86	9.82	2.44	1.95	2.11	1.90	2.24	1.88	1.76
50	14.27	14.24	14.21	2.59	1.76	2.01	1.85	2.87	1.90	1.71
100	63.84	27.44	27.38	1.60	0.74	1.36	1.25	3.14	2.51	1.16
300	405.21	105.99	105.71	1.60	0.64	1.50	1.32	12.41	2.69	1.22
1000	745,74	98.62	98.36	0.85	0.65	1.19	0.90	1.85	1.85	0.84

Table 4 Values of smoothing parameter for samples from population P4

Sample size	Histogram bin width				Kernel smoothing parameter					
	$h_{H,St}$	$h_{H,ScI}$	$h_{H,ScII}$	$h_{H,FD}$	$h_{K,IT}$	$h_{K,DPI}$	$h_{K,MS}$	$h_{K,BCV}$	$h_{K,LSCV}$	$h_{K,RR}$
30	9.17	12.14	12.11	1.08	0.18	0.63	0.72	1.80	1.42	0.66
50	163.96	192.59	192.08	2.32	0.20	1.27	4.56	21.69	3.22	8.44
100	7026.27	4049.50	4038.86	5.87	0.16	3.49	4.40	534.17	11.71	4.07
300	25392,86	7074,04	7055.45	2.01	0.10	1.50	1.82	776.89	7.90	1.88
1000	4897.75	1879.61	1874.67	2.72	0.05	0.71	0.89	97.73	2.30	0.82

Table 5 Values of smoothing parameter for samples from population P5

Sample size	Histogram bin width				Kernel smoothing parameter					
	$h_{H,St}$	$h_{H,ScI}$	$h_{H,ScII}$	$h_{H,FD}$	$h_{K,IT}$	$h_{K,DPI}$	$h_{K,MS}$	$h_{K,BCV}$	$h_{K,LSCV}$	$h_{K,RR}$
30	2840.64	3441.08	3432.03	35.29	5.13	21.12	22.55	556.49	47.9	20.87
50	374.24	446.43	446.26	12.41	4.37	12.66	14.80	49.62	27.10	13.71
100	12295.48	7150.39	7131.60	18.06	1.53	10.81	12,86	934.83	30.75	11.92
300	18352.25	5132.49	5119.00	13.51	0.68	9.60	12.15	561.61	27.25	11.25
1000	185764.00	23446.09	23384.46	8.26	0.50	7.14	8.75	2025.96	28.59	8.10

Table 6 Values of smoothing parameter for samples from population P6

For samples from normally distributed populations it can be noticed that for bigger scale parameter in populations the values of bin widths and smoothing parameters take bigger values. Moreover, they are similar

for histogram as well for kernel density estimator, especially for big sample sizes. It means that for large samples it is possible to use histogram bin width in kernel density estimation. Nevertheless all considered methods for histogram bin width are the normal reference rules and some of them are based on mean integrated squared error only the Freedman and Diaconis' rule guarantees the values of bin width close to kernel smoothing parameter. For samples from Cauchy distributed populations (with smaller number of extreme values in comparison with populations P1 and P2) histogram bin widths are significantly bigger then kernel smoothing parameters. For this type of distributions it is not recommended using histogram parameter in kernel density. For the last group of populations it can be noticed that when nonnegative random variables are regarded and distributions are asymmetric (Lévy distributed population) histogram bin width differ largely from kernel smoothing parameter (with the exception of biased cross-validation method which must have similar properties as histogram selection methods). It is caused by the difference between normal distribution (which is used in histogram rules) and asymmetric distributions such as in populations P5 and P6.

Method of choice the optimal values of kernel smoothing parameter based on power mean can be applied. Special case of power mean (for -1 as parameter of power mean; 2 as number of elements; 0.5 as generalized mean parameter) implemented for two chosen values: Freedman and Diaconis' bin width and Reference Rule smoothing parameter. Kernel density estimators for chosen samples are presented in Figures 1-4 (Gaussian kernel function and smoothing parameter calculated as power mean of histogram bin width and kernel smoothing parameter).

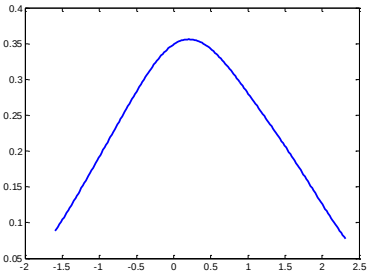


Figure 1 Kernel density estimator for sample from P1, n=30

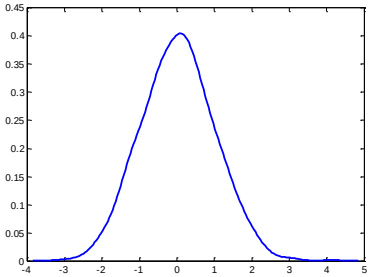


Figure 2 Kernel density estimator for sample from P1, n=1000

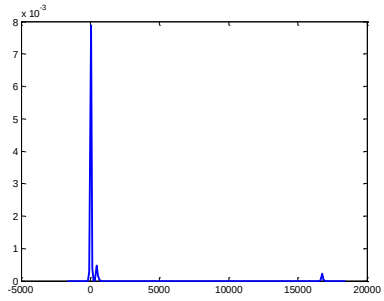


Figure 3 Kernel density estimator for sample from P6, n=30

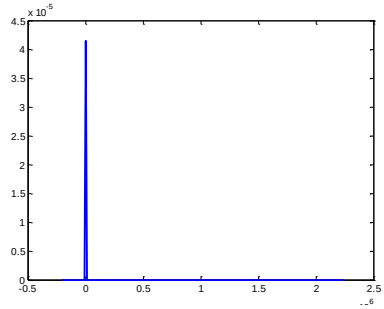


Figure 4 Kernel density estimator for sample from P6, n=1000

4 Conclusions

Methods of selection the histogram bin width are often used in practice, they are widely implemented in statistical software. Because they are based on very similar assumptions as in the case of kernel Q&D smoothing parameter's methods, there is a need to examine if these simply histogram methods can be applied in kernel density estimation.

In the case of histogram bin width, the mean integrated squared error (MISE) is the main criteria of optimality. Minimization of MISE with the assumption that data are normally distributed give simply rules for constructing the histogram. Sensitivity of MISE to bin width is broadly discussed in [10]. The possibility of using the histogram's rules in kernel density estimation is evaluated only by the comparing the values of histo-

gram bin widths and kernel smoothing parameters. The similar values of them indicate that these methods can be used interchangeably.

The following remarks should be regarded in practical applications of nonparametric density estimation:

1. it is possible to use selection methods of histogram bin width in kernel density estimation but only in the case of large samples – histogram bin widths and kernel smoothing parameters are of the similar values,
2. regarded histogram methods can be used in kernel density estimation for smaller samples when the researcher has additional information about distribution of the population - when the population distribution is close to normal distribution,
3. values of histogram bin width calculated by Freedman and Diaconis' rule are very similar to values of Reference Rule – it can indicate that these two methods are of the same properties,
4. it is possible to combine histogram and kernel selection methods in new proposed method of applying power mean of two values of regarded methods, one from histogram and one from kernel selection method. Because selection methods for histogram bin width and kernel smoothing parameter give only the estimate of optimal value, this new approach can be treated as new method in examining optimal smoothing parameter.

Further analysis should be devoted to the comparison of values of histogram bin width and kernel smoothing parameter but in the case when others kernel function in kernel density estimation are used. Especially when bounded kernel function is used what has special meaning when the random variable is bounded (for example takes only nonnegative values). Very important issue is connected with searching new method for combining histogram and kernel methods.

References

[1] Baszczyńska, A. (2016). *Smoothing Parameter in Kernel Density Estimation for Random Variables in Economic Researches* (in Polish) (First edition). Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

[2] Härdle, W. (1991). *Smooothing Techniques. With implementation in S* (First edition). New York Berlin Heidelberg London: Springer-Verlag.

[3] Devroye, L. & Lugosi, G. (2001). *Combinatorial Methods in Density Estimation* (First edition). New York: Springer-Verlag.

[4] Devroye, L. & Lugosi, G. (2004). Bin Width Selection in Multivariate Histograms by the Combinatorial Method. *Sociedad de Estadística e Investigación Operativa Test, 13* (1), 129-145.

[5] Ghosh, S. (2018). *Kernel Smoothing. Principles, Methods and Applications* (First edition). Hoboken, New Jersey: John Wiley & Sons Ltd.

[6] Gramacki, A. (2018). *Nonparametric Kernel Density Estimation and Its Computational Aspects* (First edition). Studies in Big Data, Volume 37. Springer International Publishing AG 2018.

[7] Klemelä, J. (2009). *Smoothing of Multivariate Data. Density Estimation and Visualization* (First edition). Hoboken, New Jersey: John Wiley & Sons, Inc.

[8] Knuth, K. (2013). Optimal Data-Based Binning for Histogram. https://arxiv.org/pdf/physics/0605197.pdf [05.04.2018].

[9] Lolla, S. & Hoberock, L. (2011). On Selecting the Number of Bins for a Histogram. http://worldcomp-proceedings.com/proc/p2011/DMI2080.pdf [05.04.2018].

[10] Scott, D. W. (2015). *Multivariate Density Estimation* (Second edition). Hoboken, New Jersey: John Wiley & Sons, Inc.

[11] Scott, D. W. (1979). On Optimal and Data-Based Histograms. *Biometrika, 66*(3), 605-610.

[12] Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis* (First edition). London: Chapman and Hall.

[13] Simonoff, J. (1996). *Smooothing Methods in Statistics* (First edition). New York Berlin Heidelberg: Springer-Verlag.

[14] Wand, M. & Jones, C. (1995). *Kernel Smoothing* (First edition). London: Chapman and Hall.

[15] Wang, X. & Zhang, J. (2012). Histogram-kernel Error and Its Application for Bin width Selection in Histograms. *Acta Mathematicae Applicatae Sinica,* English Series *28* (3), 607-624.

Labour market frictions and growth in small open economy: DSGE model

Jakub Bechný[1], Osvald Vašíček[2]

Abstract. In this paper, we quantify impacts of labour market disturbances on the Czech economic growth. The analysis is done by using a small open economy dynamic stochastic general equilibrium model with real wage rigidity and involuntary unemployment. To estimate the model, we use thirteen observed variables from 2001Q1 to 2017Q4 and a standard Bayesian approach. The data are not pre-filtered, but we employ the growth rates of output (and its components), which allows us to analyse impacts of various disturbances on the growth. Impulse response functions and historical shock decompositions reveal that especially labour supply shocks have had a non-negligible impact on the Czech economic growth.

Keywords: DSGE model, economic growth, labour market, small open economy

JEL classification: C32,E17
AMS classification: 91B51

1 Introduction

The Czech economy went recently through a turbulent development. The Great Recession and the following 2012-2013 crisis brought 5.75% slump of the real output,[3] and the unemployment rates of 7.8%. A risk of deflation combined with the zero lower bound then forced the Czech National Bank to adopt the exchange rate commitment during the 2013-2017 period. But at least since 2014, the Czech economy is doing well. The average output growth of 3.7% has been followed by a remarkable development in the labour market. In fact, the economy is getting very close to the full employment. By the end of the 2017, there were only 131 000 of unemployed people, roughly 236 000 of vacancies, and the unemployment rate has reached the historically lowest level of 2.5%.

Motivated by the recent development in the labour market, the aim of this paper is to quantify the impacts of the labour market disturbances on the Czech economic growth. The analysis is done by using an estimated small open economy dynamic stochastic general equilibrium (DSGE) model with labour market rigidities and involuntary unemployment. In particular, we use a modified version of a model originally proposed by Sheen & Wang [15]. The model is estimated by using the Bayesian approach and the growth rates of the real output, consumption, investment, exports, imports and wages among observed variables. This approach allows for identification of disturbances that drive economic growth.

In order to answer our research question, we analyse behaviour of the model economy in response to labour market shocks, and discuss determinants of selected variables by means of the historical shock decompositions. Our results point to a moderate impact of the labour market disturbances on the Czech economic growth. For example, the orthogonal labour supply shock reduces the unemployment rate by 0.24 percentage points, and increases growth of the real output by 0.27 percentage points (quarter-on-quarter, annualized). Furthermore, we show by means of the historical shock decompositions that the labour market shocks have had quite strong impact on the whole economy especially during the 2004-2006 time period, and from 2014 onwards.

In recent years, there has been limited research on the Czech labour market by using DSGE approach. The exceptions are Němec [13], Antosiewicz & Lewandowski [2], Pápai [14], and Tonner et al. [16]. We contribute to this literature with analysis of behaviour of the model economy in response to various labour market shocks, focused on implications for the growth. The remainder of this short paper is organized as follows. Section 2 sketches structure of our model, and Section 3 then presents results of our analysis.

2 Model

We use our modification of a small open economy model with labour market rigidities, originally proposed by Sheen & Wang [15]. These authors add more elaborate labour market block to the closed economy setting of

[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

[3]Relative to the pre-crisis peak in 2008Q3.

Christiano et. al [6], enriched with the small open economy features of Adolfson et al. [1]. All real variables of the model share a common trend due to a unit-root technology shock. The model contains four types of representative intermediate firms (domestic goods producers, consumption importers, investment importers, exporters). Capital and labour services are used only by domestic intermediate goods producers, in proportion to the relative wage and rental rate of capital. Each firm faces the Calvo [5] type of nominal rigidity, and sets its price based on markup over real marginal costs. There are thus four New Keynesian Phillips curves, and sticky prices of exporters and importers also allow for incomplete exchange rate pass-through.

The representative household attains utility form leisure, real balances and from a basket of domestically produced and imported consumption goods, subject to habit formation. The representative household also owns and accumulates the physical capital stock. The capital is accumulated by using a basket of domestically produced and imported investment goods, and the capital utilization rate can be varied at cost. Households' investment decision between domestic and foreign bonds yields uncovered interest rate parity (UIP) condition with risk premium. Behaviour of the central bank is characterized by the interest rate rule, which responds to the inflation rate, the output, and the real exchange rate. Domestic economy is assumed to be small, and domestic developments thus do not affect the rest of the world economy (which is proxied by the euro area data).

We introduced several modifications of the Sheen & Wang [15] model (focused on the foreign economy bloc, UIP condition, and on the inflation target), which were in detail discussed in Bechný & Vašíček [3]. In that paper, it was also shown that the labour market conceptualized by means of the real wage rigidity with hiring costs provides the best fit of the Czech data.[4] This variant of the labour market block will be therefore used also in this paper.

The wage setting in our model is described by means of the Nash bargaining between households and firms, who take into account the hiring costs. Following Hall [10], the real wage rigidity is constructed as the weighted average of the lagged real wage w_{t-1} and the equilibrium Nash bargaining wage w_t^*

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

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

$$w_t^* = g_t + \frac{N_t^{\sigma_L}}{\zeta_t^N \psi_t^z} - (1-\delta)\beta E_t \left[\frac{\psi_{t+1}^z}{\psi_t^z \mu_{t+1}^z \pi_{t+1}^d} (1 - x_{t+1}) g_{t+1} \right] \tag{2}$$

where g_t represents hiring costs, ζ_t^N is a stationary AR(1) labour supply preference shock, N_t is employment, ψ_t^z is households' 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 domestic inflation rate, x_t is labour market tightness (defined as the ratio of the number of hires to the number of unemployed before hiring proceeds), σ_L represents the inverse Frisch elasticity of labour supply, δ denotes average job separation rate, and β is households' discount factor. The real cost of hiring a new worker g_t, which is paid by domestic goods producers, is given by

$$g_t = \epsilon_t B x_t^{\vartheta} \zeta_t^x \tag{3}$$

where ϵ_t is a temporary technology shock, ζ_t^x is a stationary AR(1) hiring cost shock, B is scaling parameter and ϑ measures sensitivity of hiring cost to labour market tightness. Blanchard & Galí [4] show that this form of hiring cost function can be derived from a standard matching function of Mortensen & Pissarides [12].

Our model contains sixteen different structural shocks (including two labour market shocks ζ_t^N and ζ_t^x), allowing for quite detailed causal explanation of the data. More detailed information about our model concept is for sake of space not presented here, and can be found in Bechný & Vašíček [3] and in Sheen & Wang [15].

3 Empirical analysis

We stationarized all equilibrium conditions in our model by a common permanent technology process, log-linearized all model equations, and solved the whole system of sixty four equations numerically in Dynare.[5] The solved model was cast into a state-space form, and its parameters were estimated by using a standard Markov chain Monte Carlo Metropolis-Hastings algorithm and Kalman filter, as implemented in Dynare. We employed quarterly data for the Czech economy from 2001Q1 to 2017Q4. All data were taken from the Czech National Bank ARAD database, except estimates of the non-accelerating-inflation rate of unemployment (NAIRU) which were taken from the OECD database. We used the following set of thirteen observed variables: the real output,

[4] We compared four variants of the labour market block - with either real or nominal wage rigidity, and with or without hiring costs.

[5] Dynare is Matlab toolbox developed to solove, simulate and estimate DSGE models, for details see Griffoli [9].

consumption, investment, imports and exports, the real wages, CPI inflation rate, nominal interest rate (PRIBOR), the CNB's inflation target, CZK/EUR nominal exchange rate, and unemployment rate. The foreign economy block is proxied by the euro area real output, CPI inflation, and nominal interest rate (EURIBOR).

Importantly, the data were not detrended by some ad-hoc procedure like the Hodrick-Prescott [11] filtration, but we employ the growth rates of output (and of its components) and of wages as observed variables. This allows for direct analysis of the economic growth. Following Christiano et al. [7] we stationarize the growth rates, interest rates and inflation rates by demeaning.[6] And following Elbourne et al. [8] we construct stationarized measure of unemployment as deviation of actual unemployment rate from the NAIRU. We do not report results of the estimation for sake of space.[7]

3.1 Labour market and economic growth

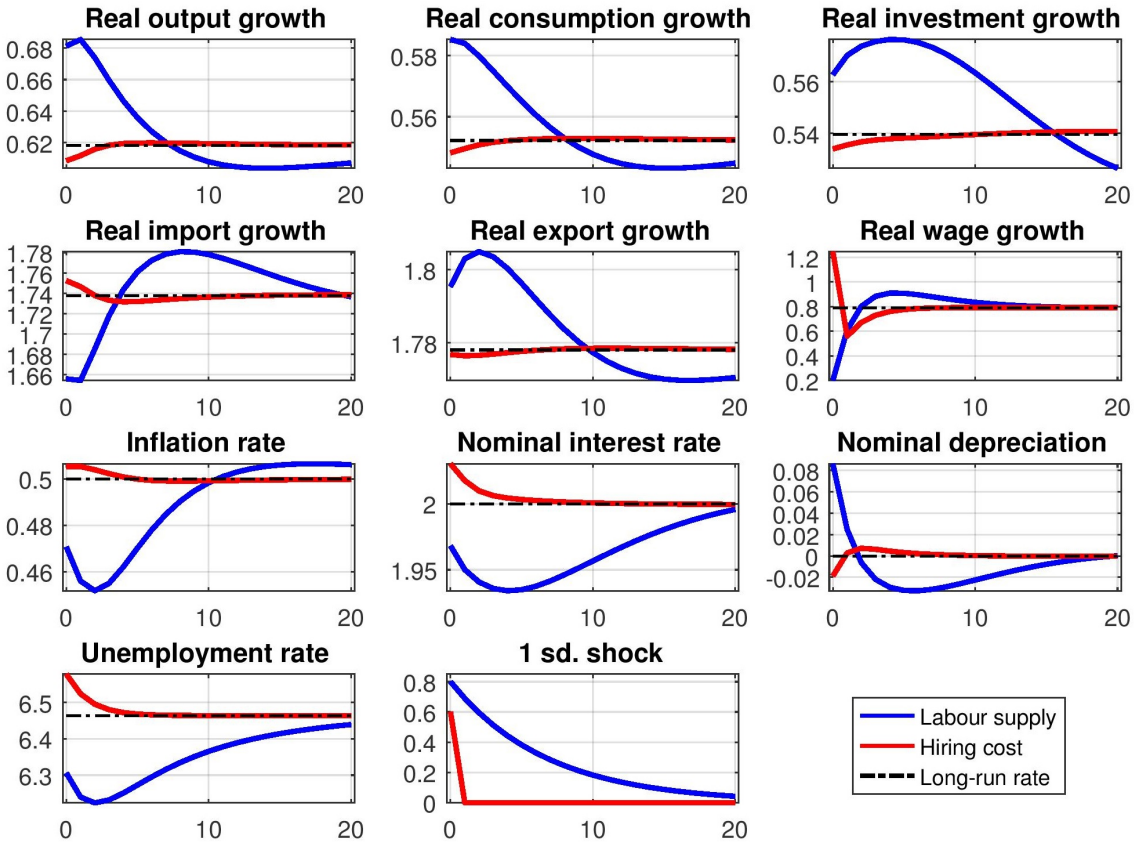


Figure 1 IRFs of observed variables to labour market shocks (QoQ growth rates in percentage points)

In this crucial section of the paper, we quantify impacts of the labour market shocks on the economic growth and discuss which shocks account for observed fluctuations in selected variables. Let us start with analysis of the impulse responses of the domestic observed variables to the orthogonal labour market shocks, shown in Figure 1. The growth rates and the inflation rate in the figure are expressed in the quarter-on-quarter units. All IRFs are rescaled such that they converge to a long-run averages of the data, e.g. on average the real output growths by 0.62 percentage points per quarter, the inflation fluctuates around 0.5%, and there is unemployment rate of 6.46%.

The one-standard-deviation labour supply shock decreases the unemployment rate by 0.24 percentage points. Increased labour supply initially lowers growth of the real wages by 0.6 percentage points, which reduces firms' marginal costs and induces (quite small) decline of the inflation, almost by 0.05 percentage points. In reaction to the inflation the central bank slightly reduces the interest rate, which is on the other hand sufficient to induce nominal depreciation of 0.09%. Increased labour supply and weaker currency have a positive effect on all sectors of the domestic economy - growth of the exports increases by 0.025 percentage points, and growth of the consumption

[6] More precisely, in order to account for possible structural change in export and import data after joining the European Union, we use separate means for the data before 2004Q3, and onwards.

[7] In total we estimated 61 parameters of our model (including persistences and volatilities of shocks, and the measurement errors). 23 parameters of the model were calibrated, usually to match the sample means of the data.

and investments increase by 0.035 percentage points. As a result, the decrease of the unemployment rate by 0.24 percentage points induces increase in the quarter-on-quarter real output growth of 0.067 percentage points (or of 0.268 percentage points in annualized terms). These results thus show that labour market shocks may potentially have a non-negligible impacts on the growth of the Czech economy.

On the other hand, the hiring cost shock has rather limited impact on the economic growth. It significantly affects only the labour market variables. The one-standard-deviation i.i.d. hiring cost shock initially induces increase of the quarter-on-quarter real wage growth by 0.45 percentage points. This is because higher hiring costs imply that firms can fill jobs more expensively, and are therefore willing to pay a higher wages to existing employees, see Equation (2). But as firms simultaneously reduce hiring of new employees, the unemployment rate increases by 0.11 percentage points which cuts growth of the real wages in the subsequent periods. As a result, the hiring cost shock reduces the quarter-on-quarter real output growth only by 0.01 percentage point.

Figure 2 shows historical shock decompositions[8] of the growth of the real output, consumption, the inflation rate (all variables are in quarter-on-quarter terms and demeaned), and of the unemployment rate (measured as deviation from the NAIRU). Overall, the dynamics of the real output growth is driven by the technology and markup shocks (with dominant role of domestic and export markups that reflect cost shocks to domestic producing and exporting firms). Consumption, which is a dominant component of the output, is driven mainly by the consumption preference shock. Inflation rate dynamics is to a large extent determined by the markup shocks (with dominant role of domestic and import markups), but an important role is played also by monetary and exchange rate UIP shocks.

Dynamics of the unemployment rate is driven mainly by the labour market shocks (i.e. the labour supply shock and hiring cost shock), but important role is played also by the technology shocks which affect productivity of labour and thus firms' labour demand as well. Increase in labour demand due to positive technology shocks has been apparent especially during the last two years. The labour market shocks had an adverse impact on the unemployment during the whole 2004-2014 period. Our estimates show that these adverse labour market shocks had a negative impact on the real output growth (and similarly on consumption) especially during the 2004Q2-2006Q1 period, between 0.2-0.3 percentage points per quarter. During approximately the same time period the labour market shocks induced the inflationary pressures of 0.12-0.23 percentage points per quarter.

From 2014 onwards, the Czech economy has been hit by a sequence of positive labour market shocks, which have, together with the technology shocks, brought the Czech unemployment almost 2 percentage points below the NAIRU. Between 2014Q1-2016Q3, the labour market shocks had a positive impact on the output growth of 0.1-0.2 percentage points per quarter. A similar quantitative effect can be observed also for the growth of consumption. On the other hand, these labour market shocks have induced disinflationary pressures of 0.1-0.17 percentage points per quarter. Impact of the labour market shocks on output growth have shrunk from 2016Q4 onwards.

Note also that the 2008-2009 Great Recession is explained mainly by the technology and markup shocks in our model (with dominant role of export and import markups). The foreign shocks had also a non-negligible impact on the economic growth. In contrast, the 2012-2013 crisis was associated with a negative consumption and markup shocks (with dominant role of domestic markup). According to our opinion, the model provides quite intuitive explanation of these events. The Great Recession is largely explained by the shocks related to the foreign development (with the permanent technology shock absorbing the simultaneous slump in all real variables). A negative consumption and domestic markup shocks during the 2012-2013 period then point to the domestic sources of the crisis, and can be attributed to the restrictive fiscal policy of the former government and its impacts on the sentiment of consumers and firms.

4 Conclusion

The goal of this short paper was to quantify the impacts of labour market disturbances on the Czech economic growth. To do so, we estimated a small open economy DSGE model with real wage rigidity and involuntary unemployment using the growth rates of observed data. We showed that the orthogonal labour supply shock has a moderate impact on the whole economy, increasing growth of the real output by 0.067 percentage points per quarter. On the other hand, the orthogonal hiring cost shock significantly affects primarily variables from the labour market, and has only a negligible impact on the overall economic growth of 0.01 percentage point.

[8]The structural shocks are aggregated into the following seven groups: Technology shocks (consisting of the permanent technology shock, temporary technology shock, and investment specific shock), Markup supply shocks (domestic markup, imported consumption markup, imported investment markup, export markup), Labour market shocks (labour supply, hiring cost shocks), Monetary policy shocks (monetary policy and inflation target shock), UIP exchange rate shock (risk premium shock to UIP condition), Consumption demand shock (consumption preference shock), and Foreign economy shocks (asymmetric foreign technology, foreign demand, supply, and monetary policy shocks).

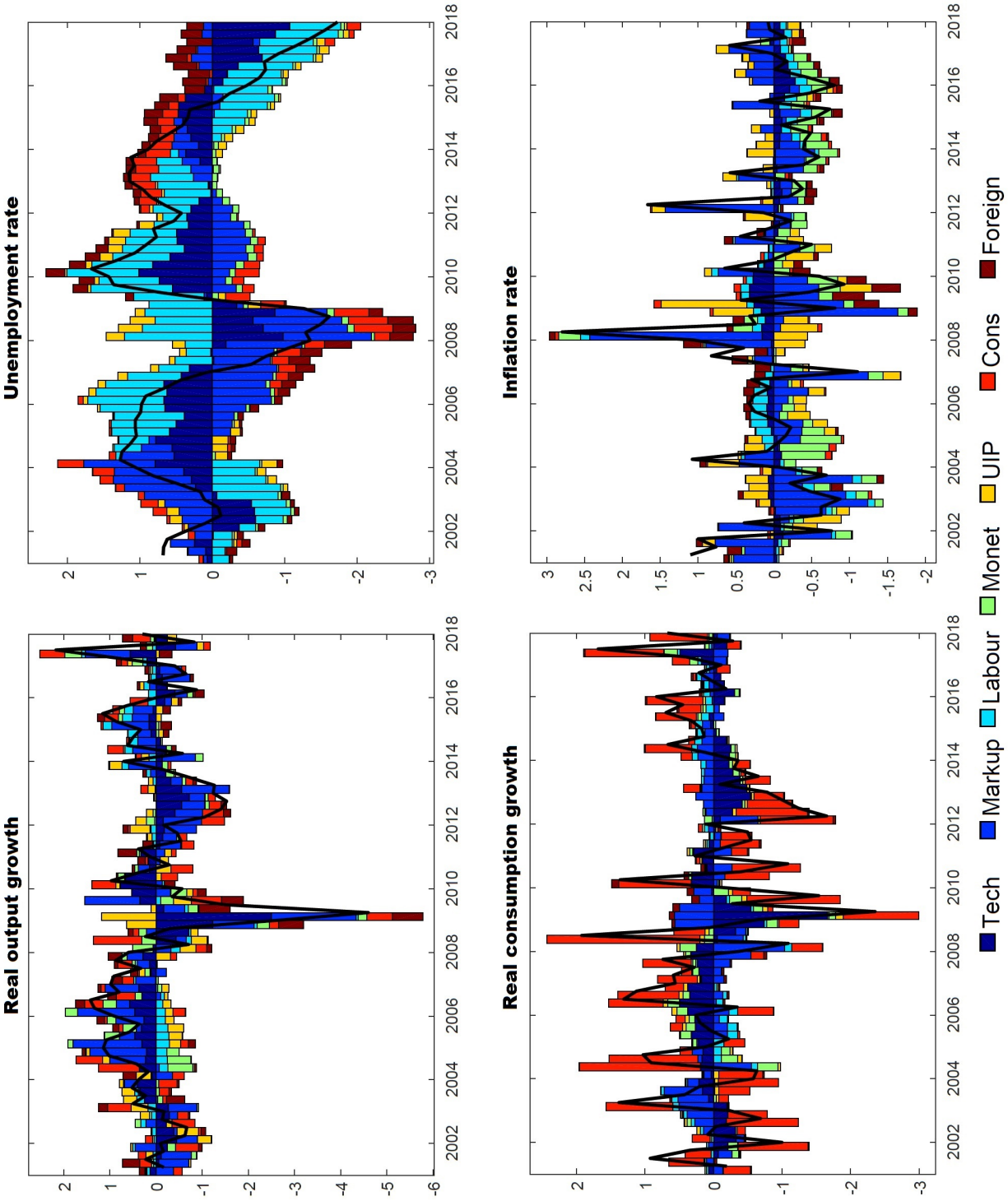


Figure 2 Historical shock decompositions of observed variables

Furthermore, we showed that dynamics of the Czech unemployment rate is driven mainly by the labour market shocks and, especially during the last two years, also by the technology shocks. We identified two time periods during which the labour market shocks had quite strong cumulative impact on the economic growth. From 2004 to 2006, the labour market shocks had an adverse effect on the output growth and induced inflationary pressures. From 2014 onwards, the labour market shocks have contributed to the output growth and induced disinflationary pressures.

Our results imply that labour market shocks may potentially have a non-negligible impacts on the development of the Czech economy. The Czech labour market is getting very close to full employment, with the historically lowest unemployment rates reaching 2.5% by the end of 2017. For many firms it is easier to bring in foreign workers to fill a labour shortage, that is likely becoming a barrier to the economic growth. Further decline in the unemployment rate is very unlikely in the Czech economy, and investigation of implications of labour shortage for economic growth and stability is thus of great importance. For further research, we would like to incorporate a permanent labour supply growth shock and non-constant labour force into the model, which will allow us to directly analyse these questions.

Acknowledgements

We would like to thank Stanislav Tvrz for helpful suggestions and comments. This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0966/2017. This support is gratefully acknowledged.

References

[1] Adolfson, S., Laséen, J., Lindé, J. & Villani, M. (2007). Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics*, 72(2), 481–511.

[2] Antosiewicz, M., Lewandowski, P. (2014). *What if you were German? DSGE approach to the Great Recession on labour markets*. IBS Working Paper 01/2014.

[3] Bechný, J. & Vašíček, O. (2017). A Medium-Scale DSGE Model with Labour Market Frictions. In P. Pražák (Ed.), *35th International Conference Mathematical Methods in Economics, Conference Proceedings* (pp. 42–47). Hradec Králové: University of Hradec Králové.

[4] Blanchard, O. J., J. & Galí, J. (2010). Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment. *American Economic Journal: Macroeconomics*, 2(2), 1–30.

[5] Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3), 383–398.

[6] Christiano, L. J., Eichenbaum, M. & Evans, Ch. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1), 1–45.

[7] Christiano, L. J., Trabandt, M. & Walentin, K. (2018). Introducing financial frictions and unemployment into a small open economy model. *Journal of Economic Dynamics and Control*, 35(12), 1999–2041.

[8] Elbourne, A., Luginbuhl, R. & de Wind, J. (2015). An estimated DSGE model for the Netherlands with unemployment. *CPB Background Document, Netherlands Bureau for Economic Policy Analysis*.

[9] Griffoli, T. M. (2013) *Dynare user guide*. Unpublished manuscript. Available from: `http://www.dynare.org/documentation-and-support/user-guide/Dynare-UserGuide-WebBeta.pdf`.

[10] Hall, R. E. (2005). Employment fluctuations with equilibrium wage stickiness. *American Economic Review*, 95(1), 50–65.

[11] Hodrick, R. J. & Prescott, E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1), 1–16.

[12] Mortensen, D. T. & Pissarides, Ch. A. (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies*, 61(3), 397–415.

[13] Němec, D. (2013) Investigating Differences between the Czech and Slovak Labour Market Using a Small DSGE Model with Search and Matching Frictions. *Czech Economic Review*, 7(1), 21–42.

[14] Pápai, A. (2017). *Modelling labour market rigidities*. Dissertation thesis. Masaryk University, Faculty of Economics and Administration, Brno. Available from: `https://is.muni.cz/th/322878/esf_d/`.

[15] Sheen, J. & Wang, B. Z. (2016). Assessing labor market frictions in a small open economy. *Journal of Macroeconomics*, 48, 231–251.

[16] Tonner, J., Tvrz, S. & Vašíček, O. (2015). *Labour Market Modelling within a DSGE Approach*. CNB Working Paper Series, 6/2015.

How wholesale fuel prices are shaped - an examination of an IPP pricing

Sylwester Bejger [1]Witold Orzeszko[2]

Abstract. The study examines time series of wholesale prices of liquid fuels (gasoline and diesel motor oil) in order to assess the mechanism of creating daily prices' levels. We use individual, daily price data of two major players in the Polish domestic market. Taking into account implied pricing mechanism existing in the market, an Import Parity Pricing schema, we examine if both of the duopolistic players can apply such mechanism. The study applies nonlinear autoregressive distributed lag (NARDL) model specification as empirical hypothesis of pricing mechanism to account for a short and long run cointegrating relationship between wholesale price and its drivers. Nonlinearities in relationships accounts for asymmetric pass through, common in a supply chain of liquid fuels. As a results, we propose an implied structure of the IPP mechanism common for both of the players by indicating the probable reference (benchmark) prices and assessing the influence of PLN/USD and PLN/EUR exchange rates. We find significant long-run relationships for both players and all of the proposed wholesale prices' drivers. We reveal that both exchange rates are important in a long term but only PLN/USD exchange rate enters implied daily prices' creation mechanism. Additionally, we compare dynamic adjustment paths of prices for the two players and show common patterns in reactions to changes in prices' determinants.

Keywords: dynamic econometric models, liquid fuel market, strategic pricing.

JEL Classification: L1 L7 C22
AMS Classification: 91B84

1 Introduction

The research examines strategic price interactions in the Polish wholesale motor fuel market in the context of determination of the most important price determinants that could enter an implied Import Parity Pricing (IPP in short) pricing schema. In a paper [4] the key characteristics of the refining industry, the wholesale market, the major players and the price's creation mechanism have been isolated and studied. Polish liquid fuel market at wholesale level is a duopolistic market with two major players: Orlen Group (PKN Orlen, or PKN) and LOTOS Group (LOTOS). The refining industry and wholesale level of a market is highly concentrated. One of the findings was that the pricing mechanism of the players corresponds to the IPP formula. It should be noted that wholesale prices of the main fuel products are publically announced on regular basis by the two major players. The prices are updated daily and announced near simultaneously about midnight on the players' web sites. As there are no official players' statements on pricing formula they utilize, the task of revealing the most important determinants entering daily prices' creation mechanism, corresponding with the IPP formula seems to be meaningful and important in both prescriptive (prediction of prices) and descriptive (analysis of competition) applications.

The IPP pricing is based on the assumption that fuel for road use is a tradable good and the ex-refinery price depends indirectly on the price of crude and of the costs of refining at domestic refineries, but rather on the price that the purchaser has to pay for this product in a relevant hub plus transport costs and other relevant spreads for the site chosen for storage. Theoretically, the IPP is the maximum level that the domestic producers' wholesale price can reach if there are no barriers to import. The IPP pricing is not an abuse of competition policy *per se*, but could foster anticompetitive practices [7]. We can point to at least two examples where the IPP pricing is officially described as domestic refineries' pricing schema, namely Portugal [19] and Australia [2]. The DAF/COMP report [9] raises the IPP policy issues as well.

2 Theoretical pricing mechanism and data description

The IPP pricing formula is basically similar across different countries and industries. In order to define the IPP policy, one has to specify two key components – the formula itself and the reference price indices. For refining

[1] Nicolaus Copernicus University, Gagarina 13a, Toruń, Poland, e-mail: sylwester.bejger@umk.pl.
[2] Nicolaus Copernicus University, Gagarina 13a, Toruń, Poland, e-mail: witold.orzeszko@umk.pl.

industry, as in [2], the formula describes wholesale net price of fuel product as a function of reference prices and some other costs:

IPP based domestic refinery price = benchmark (reference) prices + quality premium + shipping costs + wharfage + insurance and loss

Our study is focused on the most important part of the IPP policy definition - a set of benchmark price indices. In that context we want to answer the following research questions: is there cointegration between dependent price and benchmarks, which reference indices (from a commonly known set) really shape wholesale daily prices of refining products, are there differences in players' pricing policies. Choosing the candidates benchmark prices we followed a very reach literature on refining industry in different countries ([15], [10], [11] for example) and much smaller set of studies on Polish refining products' market ([6], [14], [13]). The main research hypothesis considers pricing model of a following, general form:

wholesale price of product for player i = f(exchange rate *r* (first benchmark); quotation of crude oil, petroleum product or future contract *x* (second benchmark), where i ∈ {LOTOS, PKN}.

Then we selected a set of dependent and independent variables. After necessary preprocessing, the sets contains following elements (time series): **dependent variables** (domestic products): LW_P_N (log of wholesale price of unleaded standard 95 octane gasoline reported by PKN in PLN/m3, without excise and fuel duty), LW_L_N (log of wholesale price of unleaded standard 95 octane gasoline reported by LOTOS in PLN/m3, without excise and fuel duty), LP_ON (log of wholesale price of standard diesel motor oil reported by PKN in PLN/m3, without excise and fuel duty), LL_ON (log of wholesale price of standard diesel motor oil reported by LOTOS in PLN/m3, without excise and fuel duty); **first benchmark** – exchange rates *r*: LUS_X (log of PLN/USD average exchange rate), LEUR_X (log of PLN/EUR average exchange rate); **second benchmark** – quotations *x* LB_S (log of Brent crude oil spot price in USD /m3), LNYH_S (log of New York Harbor Regular Gasoline spot price in USD/m3), LF_1 (log of New York Harbor Reformulated RBOB Regular Gasoline Future Contract 1, in USD per m3), LD_S (log of New York Harbor Ultra-Low Sulfur No 2 Diesel spot price, USD per m3). Sample period is 01.01.2006 – 31.12.2016. All of the series are transformed to 5 day's (Mon - Fri) working week.

General testable pricing models have the form (according to the types of products and benchmarks):

LW_P_N or LW_L_N = f(*r*, *x*), where *r* is LUS_X or LEUR_X and *x* is one of the regressors LB_S, LNYH_S, LF_1;

LP_ON or LL_ON = f(*r*, *x*), where *r* is LUS_X or LEUR_X and *x* is one of the regressors LB_S, LD_S.

3 Methods and results

The IPP pricing is inherently a dynamic process, though looking for the appropriate empirical model specification we focused on dynamic models only. Moreover, refining products' markets are often characterized by asymmetric pass - through ([8], [12], or [1]) confirmed partly for Polish market either [6]. This leads to nonlinearities in processes. Additionally an empirical specification should allow for cointegration analysis and be robust to integration's order incompatibilities to some extent. The above factors decided about the choice of asymmetric cointegration approach, proposed by Shin, Yu and Greenwood-Nimmo [20]. It uses a nonlinear auto-regressive distributed-lag (NARDL) model, whose structure derives from the ARDL(*p*, *q*) model ([17], [18]):

$$y_t = \alpha_0 + \sum_{j=1}^{p} \Phi_j y_{t-j} + \sum_{j=0}^{q} \boldsymbol{\theta}_j' \boldsymbol{x}_{t-j} + \varepsilon_t \quad (1)$$

NARDL(*p*, *q*) in-levels model is a natural extension of the ARDL model (1):

$$y_t = \alpha_0 + \sum_{j=1}^{p} \Phi_j y_{t-j} + \sum_{j=0}^{q} (\boldsymbol{\theta}_j^{+\prime} \boldsymbol{x}^{+}{}_{t-j} + \boldsymbol{\theta}_j^{-\prime} \boldsymbol{x}^{-}{}_{t-j}) + \varepsilon_t \quad (2)$$

where: y_t is a scalar dependent variable, $\boldsymbol{x}_t$ is a $k \times 1$ vector of regressors decomposed as: $\boldsymbol{x}_t = \boldsymbol{x}_0 + \boldsymbol{x}_t^{+} + \boldsymbol{x}_t^{-}$, Φ_j 's are the autoregressive parameters, $\boldsymbol{\theta}_j^{+}$ and $\boldsymbol{\theta}_j^{-}$ are the asymmetric distributed lag parameters, ε_t is an iid process with zero mean and constant variance σ_ε^2. The key role in our study plays conditional error correction form (conditional ECM, [17], [20]):

$$\begin{aligned} \Delta y_t &= \rho \xi_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + \sum_{j=0}^{q-1} (\boldsymbol{\pi}_j^{+\prime} \Delta \boldsymbol{x}^{+}{}_{t-j} + \boldsymbol{\pi}_j^{-\prime} \Delta \boldsymbol{x}^{-}{}_{t-j}) + \varepsilon_t = \\ &= \rho y_{t-1} + \boldsymbol{\theta}^{+\prime} \boldsymbol{x}^{+}{}_{t-1} + \boldsymbol{\theta}^{-\prime} \boldsymbol{x}^{-}{}_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + \sum_{j=0}^{q-1} (\boldsymbol{\pi}_j^{+\prime} \Delta \boldsymbol{x}^{+}{}_{t-j} + \boldsymbol{\pi}_j^{-\prime} \Delta \boldsymbol{x}^{-}{}_{t-j}) + \varepsilon_t \end{aligned} \quad (3)$$

where: $\xi_t = y_t - \boldsymbol{\beta}^{+\prime}\boldsymbol{x}_t^{+} + \boldsymbol{\beta}^{-\prime}\boldsymbol{x}_t^{-}$ is the nonlinear error correction term, $\boldsymbol{\beta}^{+} = -\frac{\boldsymbol{\theta}^{+}}{\rho}$ and $\boldsymbol{\beta}^{-} = -\frac{\boldsymbol{\theta}^{-}}{\rho}$ are the asymmetric long-run parameters (long run multipliers), $\boldsymbol{\pi}_i^{+}$ and $\boldsymbol{\pi}_i^{-}$ are short-run (impact) parameters, and ρ is an error correction coefficient. Specification (4) allows for long and short run testing for nonlinearity and cointegration. There are two tests for the existence of a stable long-run levels relationship we apply: the t_{BDM}-statistic proposed by Banerjee et al. [3] which tests: $H_0 : \rho = 0$ (no long run level relationship) and $H_1 : \rho < 0$ and the F_{PSS} statistics by Pesaran, Shin and Smith [17], which tests: $H_0 : \rho = \boldsymbol{\theta}^{+} = \boldsymbol{\theta}^{-} = 0$ vs. $H_1 : \rho = \boldsymbol{\theta}^{+} = \boldsymbol{\theta}^{-} \neq 0$.

3.1 Empirical examination

Empirical examinations contains few steps in this case (usual preliminary steps, such as visual inspection of the series, descriptive stats, integration order testing, structural breaks testing) and final models' selection, estimation and inference. However, for the sake of compactness we report only the most important results.

For the bound testing for cointegration based on (3) the key step is an integration's order testing. This approach yields consistent estimates of the long-run coefficients that are asymptotically normal irrespective of whether the underlying regressors are I(1) or I(0) but bound test could lead to spurious results in the presence of I(2) variables. Therefore the usual unit root tests - Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) were carried out to check the stationarity and number of integration order of the variables. Both tests showed that the processes are integrated of order 1 or 0. Visual inspection showed that the series exhibit structural breaks. We confirm a break in each series by a proper test and decided to check an order of integration by Perron [16] unit root test with structural break points. The obtained results reports Table 1:

Series	Estimated break date	ADF t-Statistic	Prob.*
LW_P_N	12/29/2008	-3.896	0.1950
LW_L_N	12/30/2008	-3.707	0.2791
LB_S	9/04/2014	-2.955	0.7103
LNYH_S	9/29/2014	-3.230	0.5494
LUS_X	7/14/2014	-2.806	0.7843
LEUR_X	8/21/2008	-4.971	0.0010
LP_ON	3/18/2009	-3.101	0.6263
LL_ON	12/11/2009	-2.984	0.6953
LD_S	9/03/2014	-2.849	0.7632
LF_1	9/25/2014	-3.143	0.6008

Test critical values: -4.949(1%), -4.443 (5%), -4.193 (10%)
*Vogelsang (1993) asymptotic one-sided p-values.

Table 1 Results of a Parron test

The results in Table 1 indicate that all of the time series are I(1) or I(0) and we reject a null hypothesis of integration of higher order.

Next we move to central part of an examination, i.e. IPP model estimation. At first we want to check how the exchange rates enters the pricing mechanism. To answer that question we estimated ARDL(*p*, *q*) models (1) for every proper sets of variables. The results shows Table 2.

Depend. Var.	**LW_P_N**			**LW_P_N**			**LW_L_N**			**LW_L_N**		
Regressor *x*	**LB_S**			**LB_S**			**LNYH_S**			**LNYH_S**		
Exchange rate *r*	**LUS_X**			**LEUR_X**			**LUS_X**			**LEUR_X**		
Model estimated	**ARDL(6,6,6)**			**ARDL(6, 6, 4)**			**ARDL(7,11,6)**			**ARDL(7, 11, 5)**		
Parameter	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**
ρ	-0.019	-6.289	0.0000	-0.012	-6.244	0.0000	-0.035	-6.539	0.0000	-0.017	-6.369	0.0000
β_r	1.201	16.884	0.0000	1.765	7.703	0.0000	1.129	28.417	0.0000	1.725	12.639	0.0000
β_x	0.888	20.784	0.0000	0.817	11.005	0.0000	0.954	52.350	0.0000	0.887	18.228	0.0000
π_r	0.202	11.721	0.0000	0.015	0.600	**0.5483**	0.144	7.907	0.0000	0.004	0.151	**0.8800**
π_x	0.140	18.281	0.0000	0.119	15.224	0.0000	0.093	13.503	0.0000	0.079	10.837	0.0000
Cointegr. tests	Stat. Value			Stat. Value			Stat. Value			Stat. Value		
F_PSS	13.196			13.713			14.587			14.723		
t_BDM	-6.289			-6.244			-6.539			-6.369		
Diagnostics	Stat. Value		Prob.	Stat. Value		Prob.	Stat. Value		Prob.	Stat. Value		Prob.
BG Ser. Cor. Test	0.169		0.917	0.226		0.878	1.129		0.336	0.517		0.671
RESET Test	1.827		0.177	0.174		0.677	0.469		0.493	0.057		0.812
Adj. R-squared	**0.399**			**0.345**			**0.398**			**0.327**		

Akaike criterion	**-6.684**	**-6.598**	**-6.563**	**-6.444**

Table 2 ARDL estimation results (PLN/USD and PLN/EUR exchange rates)

Table 2 contains results representative for a whole group of models. Parameters' names are coherent with convention of equation (3), i.e. β is a long run multipliers (symmetric) and π is the parameter for a first difference, (not lagged) of a regressor. It shows estimates and statistics which are the most important for the study. We can conclude that there is significant long term relationship in each case but PLN/EUR exchange rates do not enter daily pricing mechanism. The models with PLN/EUR perform generally worse in a sense of usual goodness of fit measures and short run effects on the product prices, which are not significant. As a result of this step we decided to drop the PLN/EUR rate and focus on pricing models with PLN/USD exchange rate only. Following that we estimated 10 unrestricted NARDL (p, q_1, q_2, q_3, q_4) models, using ECM form specified in eq. (3). The results of estimation shows Table 3.

Depend. Var.	**LW_P_N**			**LW_L_N**			**LW_P_N**			**LW_L_N**			**LW_P_N**		
Regressor *x*	**LB_S**			**LB_S**			**LNYH_S**			**LNYH_S**			**LF_1**		
Exchange rate	**LUS_X**			**LUS_X**			**LUS_X**			**LUS_X**			**LUS_X**		
Model estimated	**NARDL(6,6,6,5,6)**			**NARDL(6,6,6,4,6)**			**NARDL(6,7,7,4,6)**			**NARDL(7,7,8,4,6)**			**NARDL(6,7,7,5,6)**		
Parameter	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**
ρ	-0.020	-6.49	0.000	-0.022	-6.51	0.000	-0.055	-9.59	0.000	-0.059	-9.22	0.000	-0.019	-7.29	0.000
β_r^-	1.133	7.12	0.000	1.086	7.49	0.000	0.832	13.40	0.000	0.830	13.22	0.000	1.221	6.35	0.000
β_r^+	1.155	8.22	0.000	1.117	8.74	0.000	0.884	17.94	0.000	0.884	18.03	0.000	1.173	7.51	0.000
β_x^-	0.849	12.26	0.000	0.829	11.74	0.000	0.866	33.98	0.000	0.865	34.33	0.000	1.074	12.93	0.000
β_x^+	0.838	11.28	0.000	0.814	10.84	0.000	0.849	29.04	0.000	0.848	28.89	0.000	1.099	11.33	0.000
π_r^-	0.179	5.38	0.000	0.163	4.58	0.000	0.117	3.62	0.000	0.086	2.48	0.013	0.128	3.84	0.000
π_r^+	0.223	7.51	0.000	0.177	5.58	0.000	0.208	7.20	0.000	0.171	5.52	0.000	0.226	7.59	0.000
π_x^-	0.114	8.10	0.000	0.114	7.60	0.000	0.104	9.29	0.000	0.116	9.58	0.000	0.108	7.76	0.000
π_x^+	0.168	12.20	0.000	0.140	9.52	0.000	0.083	6.90	0.000	0.078	6.05	0.000	0.108	7.32	0.000
Cointegr. tests	Stat. Value			Stat. Value			Stat. Value			Stat. Value			Stat. Value		
F_PSS	8.658			8.690			18.97			16.98			10.61		
t_BDM	-6.494			-6.511			-9.586			-9.076			-6.461		
Diagnostics	Stat.Value		Prob.	Stat. Value		Prob.	Stat.Value		Prob.	Stat. Value		Prob.	Stat. Value		Prob.
BG Ser. Cor. Test	0.480		0.696	1.719		0.161	0.881		0.450	1.682		0.169	0.473		0.701
RESET Test	1.075		0.300	0.251		0.616	0.666		0.415	0.052		0.820	3.198		0.074
Adj. R-squared	**0.401**			**0.387**			**0.418**			**0.405**			**0.400**		
Akaike criterion	**-6.68**			**-6.55**			**-6.71**			**-6.57**			**-6.68**		

Table 3 NARDL estimation results

Depend. Var.	**LW_L_N**			**LP_ON**			**LL_ON**			**LP_ON**			**LL_ON**		
Regressor *x*	**LF_1**			**LB_S**			**LB_S**			**LD_S**			**LD_S**		
Exchange rate	**LUS_X**			**LUS_X**			**LUS_X**			**LUS_X**			**LUS_X**		
Model estimated	**NARDL(6,7,7,4,6)**			**NARDL(5,8,6,6,8)**			**NARDL(2,7,6,4,6)**			**NARDL(5,7,7,6,9)**			**NARDL(2,9,7,4,6)**		
Parameter	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**	**Value**	**t-Stat.**	**Prob.**
ρ	-0.020	-6.99	0.000	-0.025	-7.43	0.000	-0.022	-7.09	0.000	-0.030	-6.85	0.000	-0.029	-7.05	0.000
β_r^-	1.206	6.27	0.000	0.636	4.58	0.000	0.595	4.03	0.000	0.804	10.02	0.000	0.810	9.34	0.000
β_r^+	1.163	7.60	0.000	0.703	5.72	0.000	0.670	5.25	0.000	0.794	10.85	0.000	0.799	10.34	0.000
β_x^-	1.070	13.41	0.000	0.828	21.81	0.000	0.807	21.23	0.000	0.829	25.93	0.000	0.825	26.49	0.000
β_x^+	1.093	11.51	0.000	0.798	18.05	0.000	0.773	16.97	0.000	0.842	23.65	0.000	0.838	23.51	0.000
π_r^-	0.108	3.04	0.002	0.170	6.05	0.000	0.150	5.62	0.000	0.151	5.36	0.000	0.125	4.70	0.000
π_r^+	0.186	5.83	0.000	0.186	7.38	0.000	0.196	8.26	0.000	0.185	7.34	0.000	0.206	8.72	0.000
π_x^-	0.119	8.00	0.000	0.126	10.40	0.000	0.117	10.25	0.000	0.109	8.47	0.000	0.117	9.63	0.000
π_x^+	0.091	5.76	0.000	0.126	10.75	0.000	0.120	10.87	0.000	0.090	7.00	0.000	0.103	8.42	0.000
Cointegr. tests	Stat. Value			Stat. Value			Stat. Value			Stat. Value			Stat. Value		
F_PSS	9.761			11.86			10.74			9.638			10.23		
t_BDM	-6.203			-7.425			-7.095			-6.847			-7.046		
Diagnostics	Stat. Value		Prob.	Stat. Value		Prob.	Stat. Value		Prob.	Stat. Value		Prob.	Stat. Value		Prob.
BG Ser. Cor. Test	1.991		0.113	0.518		0.670	1.114		0.342	1.832		0.139	0.478		0.698
RESET Test	2.101		0.147	1.150		0.284	1.114		0.291	4.658		0.031	4.974		0.026
Adj. R-squared	**0.386**			**0.418**			**0.460**			**0.414**			**0.464**		
Akaike criterion	**-6.54**			**-7.03**			**-7.13**			**-7.03**			**-7.14**		

F_PSS, t_BDM: F-statistics of F_PSS and t-statistics of t_BDM bound testing approach, the critical values for Case 3 unrestricted intercept, no trend, k = 4 and usual significance levels: F-stat.: I(0); I(1) -> 1%, 3.74; 5.06; 5%; 2.86; 4.01; 10%;2.45;3.52; t-stat.: I(0); I(1) -> 1%; -3.43; -4.6; 5%; -2.86; -3.99; 10%; -2.57; -3.66.

Table 3 NARDL estimation results (cont.)

Discussing results contained in Table 3 we have to remember that $\beta_r^+ = -\frac{\theta_r^+}{\rho}$, $\beta_r^- = -\frac{\theta_r^-}{\rho}$ are the asymmetric long run parameters for PLN/USD exchange rate, $\beta_x^+ = -\frac{\theta_x^+}{\rho}$, and $\beta_x^- = -\frac{\theta_x^-}{\rho}$ are the assymetric long-run parameters for a regressor x, $\pi_r^+, \pi_r^-, \pi_x^+, \pi_x^-$ are parameters capturing the most direct short-run impact of regressors and ρ is an error correction coefficient. In all of the models the values of statistics of the F_PSS and t_BDM tests allow us to reject the null hypothesis of no-cointegration. The values of the estimated coefficients playing a key role in a research, i.e. the error correction coefficients, the long-run parameters and the impact parameters (capturing the most direct short run transmission) are significant at 1% level. Comparing the results for two estimated ARDL models with the PLN/USD exchange rate (Table 2) and the results of Table 3 one can see that NARDL models perform generally better than ARDL in terms of goodness of fit. This is the case for the whole set of models. Estimated models exhibit no serial correlation in error terms and stability (except models with LD_S regressor). In almost all of the cases we observe that $\pi_r > \pi_x$ (regardless of the sign). It shows that PLN/USD exchange rate enters daily pricing mechanism. Analyzing a set of benchmarks for unleaded standard 95 octane gasoline we find models with LNYH_S performing the best, either in terms of goodness of fit and speed of adjustment to an equilibrium in one day. Future contract and Brent crude quotations influence the wholesale prices of gasoline in a long run and probably do not enter IPP in a daily formula. In a case of the models for diesel oil's pricing the NYH Ultra-Low Sulfur No 2 Diesel spot price seems to be better benchmark, unfortunately both models with that regressor lack of stability. Moving to strategic comparison (between players) we can observe significant differences in the short term parameters for the PLN/USD exchange rate in all of the cases. Dynamic multipliers depicted in Figure 1 shows some differences as well. The models for both players differ in terms of the statistical properties either. It drives us to the conclusion that the implied IPP formulas of both major players could contain the same benchmarks but differ on a level of a formula construction.

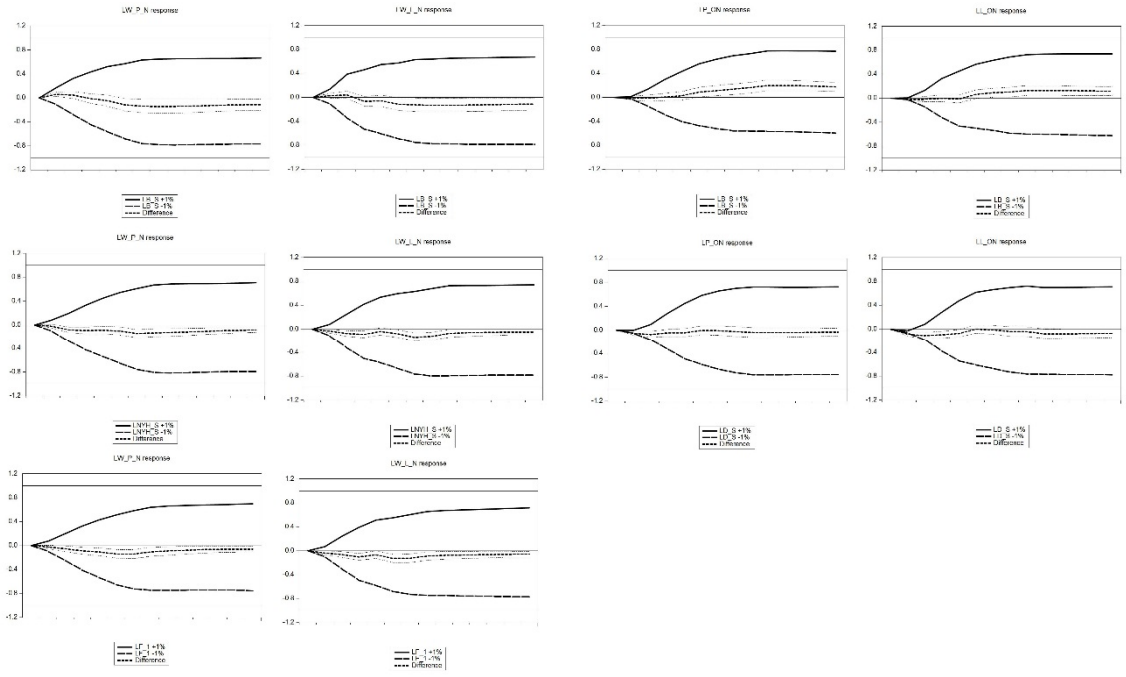


Figure 1 Dynamic multipliers

4 Conclusions

In this paper we analyzed individual daily price data of the two major players (with cumulative market share of 90%) in a Polish fuel market on a wholesale level of distribution to examine the IPP mechanism of prices' creation. Referring to the research questions formulated at the start, we can conclude that a significant long-run relationship exists in the cases of all of the examined benchmarks and the wholesale products. The NARDL models consisting dynamic long and short run asymmetry occurred to be superior to the symmetric ARDL models, in terms of the usual information criterions and goodness of fit. We showed that the PLN/USD exchange rate enters the implied daily pricing formula as a main price driver, which is coherent with the IPP hypothesis. The PLN/EUR exchange rate influences the products' prices in a long term. The IPP mechanism of price creation seems to be confirmed in

a case of regular 95 octane gasoline and the models with a NYH gasoline spot price benchmark, which occurred to be superior to the others and reveals similar dynamics for both PKN and LOTOS in transmitting the NYH gasoline price changes. This suggests that the NYH spot price or a similar quotation (maybe more appropriate for Central Europe market) could be probable IPP factor. Strategic behavior of both of the players for this benchmark can be consistent with the focal price equilibrium of a strategic game showed in Bejger [5]. We found evidences that a diesel spot price could be a better benchmark for the IPP schema than a Brent spot price in a case of motor diesel oil prices' creation, however there is strong need to use a quotation from ARA hub because of instability of estimated models. Finally, it seems that implied IPP formulas of both major players could contain the same benchmarks but differ on a level of a formula specification.

References

[1] Atil, A. Lahiani, A. & Nguyen, D.K. (2014). Asymmetric and nonlinear pass-through of crude oil prices to gasoline and natural gas prices. *Energy Policy 65*, 567–573.

[2] Australian Competition and Consumer Commission. (2007). Petrol prices and Australian consumers: *Report of the ACCC inquiry into the price of unleaded petrol*. December 2007.

[3] Banerjee, A. Dolado, J. & Mestre, R. (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis 19*, 267-283.

[4] Bejger S. (2015). Investigation of the nature of strategic interactions in the Polish wholesale fuel market: statistical analysis of a market structure and a price mechanism, *Acta Universitatis Nicolai Copernici Ekonomia, 46 (1)*, 73-95.

[5] Bejger, S., (2016). Theoretical Model of Pricing Behavior on the Polish Wholesale Fuel Market, *Folia Oeconomica Stetinensia 16*, 286-300.

[6] Bejger, S. & Bruzda, J. (2002). Identification of market power using test for asymmetric pricing – an example of Polish petrochemical industry. *Dynamic Econometric Models 5*, 135-146.

[7] Bonakele. T. (2014). Import Parity Pricing and Competition Policy. *http:\\PCTI/20140826/CoB/Competitioncommission/31*.

[8] Borenstein, S. & Shepard, A. (2002). Sticky prices, inventories, and market power in wholesale gasoline markets. *RAND Journal of Economics 33*, 116-139.

[9] DAF/COMP. (2013). Competition in road fuel – executive summary of that debate and the documents from the meeting. *Organisation for Economic Co-operation and Development*.

[10] Eckert, A. (2013). Empirical studies of gasoline retailing: a guide to the literature. *Journal of Economic Surveys 27*, No. 1, 140–166.

[11] Energy Information Administration. (2014). What Drives U.S. Gasoline Prices? *U.S. Department of Energy*.

[12] Frey, G. & Manera, M. (2007). Econometric models of asymmetric price transmission. *Journal of Economic Surv. 21*, 349–415.

[13] Leszkiewicz-Kędzior, K. (2012). Modelling Fuel Prices. An I(1) Analysis. *CEJEME 2*, 75-95.

[14] Miłobędzki, P. (2008). Orlen or Lotos? Which is Setting Prices at the Wholesale Market for Unleaded Petrol in Poland? *Dynamic Econometric Models Vol. 8*.

[15] Oladunjoye, O. (2008). Market structure and price adjustment in the U.S. wholesale gasoline markets. *Energy Economics 30*, 937–961.

[16] Perron, P. (1989). The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. *Econometrica 57*, 1361-1401.

[17] Pesaran, H.M. Shin, Y.C. & Smith, J.R. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics 16*, 289–326.

[18] Pesaran, M. H. & Shin,Y. (1999). An autoregressive distributed lag modelling approach to cointegration analysis. Chapter 11 in S. Strom (ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, Cambridge University Press, Cambridge.

[19] Portuguese Competition Authority. (2009). Detailed analysis of the liquid fuel and bottled gas sectors in Portugal Final Report. *Portuguese Competition Authority (Autoridade da Concorrência)*, March 2009.

[20] Shin, Y. Yu, B. & Greenwood-Nimmo, M.J. (2013). Modelling Asymmetric Cointegration and Dynamic Multipliers in a Nonlinear ARDL Framework. in *Festschrift in Honor of Peter Schmidt*, William C. Horrace and Robin C. Sickles (Eds.). Springer Science & Business Media: New York.

Modeling of e-shop selection processes the presence of price comparison sites

Ladislav Beranek[1], Radim Remes[2]

Abstract. A price comparison website is one of successful online business applications. They offer to users (e.g., Heureka.cz or Zbozi.cz) a possibility to provide them with an overview of prices of some product in various e-shops including reputation of these e-shops, thereby they also influence consumers' perceptions of the risks associated with online shopping. The aim of our paper is to present a model that would model the basic processes of the choice of certain e-shop if a user uses one of the price comparator site. We use Dempster-Shafer theory to the modelling of this process. Our model should predict users' selection of certain e-shop. While the product is recommended based on a recommendation system (or other, e.g., social network), e-shop is chosen based on a price of this product and reputation of this e-shop outlined in price comparison site. We use data collected from performed experiment to verify our model. It follows from a comparison that the model has good predictive abilities.

Keywords: E-commerce, Mathematical model, Price comparison site, Dempster-Shafer theory.

JEL Classification: D85, L81
AMS Classification: 90B60

1 Introduction

Simultaneously with the development of e-business, related business applications have been developed and many new approaches have been widely used as for example recommendation systems or other. Price comparisons that allow comparison of prices in different online stores are the most successful. These sites are also known as price comparisons, shopbots, or internet buying agents. Buyers online use them to obtain information about the price or the relevant stores. They reduce buyer search costs and help them make decisions by providing price comparison information that is rarely found in the context of physical retail purchases [6]. Compare pricing pages were mainly surveyed in terms of the impact of price comparison pages on the price of products and services and the sensitivity of online shoppers to the price [15, 11]. Degeratu et al. [8] states that the existence of price comparison sites increases price competition and buyer price sensitivity. We can look at e-commerce also as a network. Social and business networks are increasingly important areas of re-search in many disciplines [8, 9, 10]. However, stable equilibrium and models have focused primarily on this, while their dynamics and productivity have gained limited attention on research. One of the main tasks is to better understand, anticipate and control their dynamics, including how they form, develop and shape their behaviors and performances [1, 12, 13]. Enough progress has been made in e-commerce applications so far, and e-commerce plays a very important role in the economy. A large number of buyers and sellers cooperate with each other via web site transactions [2, 3, 5]. These interactions support development and shape complex e-commerce market structures. Getting a deep look at e-commerce research has a deep and lasting significance.

Our paper presents a model that would outline the basic processes of online shopping if the user uses one of the price comparators. The problem description is as follows. The customer is looking for a specific product (such as a robotic vacuum cleaner) he has heard from acquaintances, from social networking friends, etc. A specific product can also be recommended to him/her by a recommendation system on a website or on a specific e-shop site. The customer gets a certain idea of the products he has at his disposal. He/she may also have a preference of a certain brand, for example iRobot when he/she buy robotic vacuum cleaner. He/she wants to buy this product at the lowest price. He therefore clicks on the website of a price comparison site (Heureka.cz). He/she selects chosen product and finds a range of e-shops selling the product. Offers of e-shops are sorted by price, lowest to highest. For each e-shop, ratings of previous customers are displayed. The customer chooses now e-shop according the price and rating of this e-shop and then buys the chosen product in this e-shop (Figure 1, Phase *i*). In real life, people often resort to friends in their social networks for advice before purchasing a product or consuming a service (Figure 1, Phase *ii*). We will model the Phase *ii* (Figure 1) by the diffusion: resources are transferred from users to their

[1] University of South Bohemia, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentska 13, Ceske Budejovice, beranek@ef.jcu.cz.
[2] University of South Bohemia, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentska 13, Ceske Budejovice, inrem@ef.jcu.cz.

friends in social network and redistributed among items' neighboring users in user-e-shop bipartite network, friends in social network will influence the final selection of certain e-shop. The phase *i* is modelled with the help of belied function theory (Dempster-Shafer theory [14]), the influence of friends in social network is modelled as correction of resulting belief function.

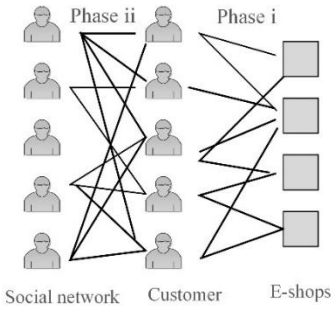


Figure 1 The e-shop selection process

The article is structured as follows: Section 2 briefly reviews the used methods with some definitions that we use throughout this article. Algorithmic details can be found in Section 3. Section 4 describes the experimental procedure and results, followed by conclusions in Section 5.

2 Methods

Basic Notions of Dempster-Shafer theory

Considering a finite set referred to as *the frame of discernment* Ω, a *basic belief assignment (BBA)* is a function $m: 2^{\Omega} \rightarrow [0,1]$ so that

$$\sum_{A \subseteq \Omega} m(A) = 1, \tag{1}$$

where $m(\emptyset) = 0$, see [14]. The subsets of 2^{Ω} which are associated with non-zero values of *m* are known as *focal elements* and the union of the focal elements is called *the core*. The value of *m(A)* expresses the proportion of all relevant and available evidence that supports the claim that a particular element of Ω belongs to the set *A* but not to a particular subset of *A*. This value pertains only to the set *A* and makes no additional claims about any subsets of *A*. We denote this value also as a *degree of belief* (or *basic belief mass - BBM*).
Shafer further defined the concepts of *belief Bel*(·) and *plausibility Pl*(·), see [14]. A *BBA* can also be viewed as determining a set of probability distributions *P* over Ω so that $Bel(A) \leq P(A) \leq Pl(A)$. It can be easily seen that these two measures are related to each other as $Pl(A) = 1 - Bel(\neg A)$. *Moreover, both of them are equivalent to m.* Thus one needs to know only one of the three functions *m*, *Bel*, or *Pl* to derive the other two. Hence we can speak about belief function using corresponding *BBA*s in fact.
Dempster's rule of combination can be used for pooling evidence represented by two belief functions Bel_1 and Bel_2 over the same frame of discernment coming from independent sources of information. The Dempster's rule of combination for combining two belief functions Bel_1 and Bel_2 defined by (equivalent to) *BBA*s m_1 and m_2 is defined as follows (the symbol ⊕ is used to denote this operation):

$$(m_1 \oplus m_2)(A) = \frac{1}{1-k} \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \tag{4}$$

where

$$k = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C) \cdot \tag{5}$$

Here *k* is frequently considered to be a *conflict measure* between two belief functions m_1 and m_2 or a measure of conflict between m_1 and m_2 [14]. Unfortunately, this interpretation of *k* is not correct, as it includes also internal conflict of individual belief functions m_1 and m_2 [7]. Demspter's rule is not defined when $k = 1$, i.e. when cores of m_1 and m_2 are disjoint. This rule is commutative and associative; as the rule serves for the cumulating of beliefs, it is not idempotent.

3 Belief Function Correction

The combination rule is usually denoted as the orthogonal sum of belief values; in other words, the combination of belief from evidence *A* and belief from evidence *B* is denoted as $Bel_{A,B} = Bel_A \oplus Bel_B$. Therefore, the global belief of *A* can be represented as $Bel(A) = \oplus Bel_i$, for all pieces of evidence that supports *A*. To illustrate the concepts, consider a subset $X \subseteq 2^{\Theta}$ and evidence E_1 that yield a set of values represented by $m_{E1}(\{x\})$, $m_{E1}(\{\neg x\})$, and $m_{E1}(\{x, \neg x\})$ (we can this last term also write down as $m_{E1}(\Theta)$ for Θ is the frame of discernment which has

two elements $\Theta = \{x, \neg x\}$). Suppose that evidence E_1 may provide, in general, some support that X is true, i.e., event x occurs, or some support that X is not true, i.e., event $\neg x$ occurs. In terms of the mass function, the *BBA*s for x and $\neg x$ are $m_{E1}(\{x\})$ and $m_{E1}(\{\neg x\})$, respectively. Lack of knowledge about whether x occurs or not is represented by $m_{E1}(\{x, \neg x\})$. The sum of the three values is one. i.e., $m_{E1}(\{x\}) + m_{E1}(\{\neg x\}) + m_{E1}(\{x, \neg x\}) = 1$. We can further assume that evidence E_1 is either reliable with probability 0.9 or unreliable with probability 0.1. Now, we can use fraudulent seller behavior as an example [4]: suppose that evidence E_1 supports that seller i sells stolen goods is with 100% certainty. Considering E_1's reliability, E_1 gives 0.9 degree of belief for supporting that bidder i is a fraudulent seller (i.e., $m_{E1}(\{x\}) = 0.9$), but zero degree of belief that seller i is honest (i.e., $m_{E1}(\{\neg x\}) = 0$) because the evidence does not support seller i is honest. The remaining degree of belief (0.1) is due to the uncertainty, i.e., $m_{E1}(\{x, \neg x\}) = 0.1$.

Belief function correction. When receiving a piece of information represented by a belief function, some metaknowledge regarding the quality or reliability of the source that provides some information can be available. In the following paragraphs, we describe briefly some possibilities how to adjust the information according to this metaknowledge.

Discounting. To handle the lower reliability of information sources, a discounting scheme has been introduced by Shafer [14]. It is expressed by equations:

$$ {}^{\alpha}m(A) = \begin{cases} (1-\alpha)m(A) & \text{if } A \subset \Omega \\ \alpha + (1-\alpha)m(\Omega) & \text{if } A = \Omega \end{cases}, \tag{6} $$

where $\alpha \in [0,1]$ is a *discounting factor* and ${}^{\alpha}m(A)$ denotes the discounted mass of $m(A)$. The larger α is, the more mass $m(A)$ is withdrawn from $A \subset \Theta$ and assigned to the frame of discernment Θ.

Thus, the principle of discounting is *transferring parts of basic belief masses BBMs* of all focal elements which are proper subsets of the frame of discernment *to the entire frame*. This process is the result of additional information which indicates that the source is not entirely reliable. The transfer of *BBM*s from a source to the framework reflects an increase of the degree of uncertainty of the data that the source produces.

De-discounting (reinforcement). In some cases we need to perform an opposite operation, e.g., transfer parts of basic belief mass (*BBM*) from the entire frame to all focal elements. This can be the result of a situation when we, for example, obtain information that the source of the information is more reliable than we had anticipated at the beginning. We can then re-compute m by reversing the discounting operation [4]. We denote this operation as *reinforcement* (or *de-discounting*):

$$ m(A) = \frac{{}^{\alpha}m(A)}{1-\alpha} \qquad \forall A \subset \Omega \tag{7} $$

$$ m(\Omega) = \frac{{}^{\alpha}m(\Omega) - \alpha}{1-\alpha} \ , \tag{8} $$

where $\alpha \in [0, {}^{\alpha}m(\Omega)]$. We denote here α as a reinforcement coefficient. The result of maximal de-discounting is the totally reinforced belief function. It is noted ${}^{tr}m$ and defined as follows:

$$ {}^{tr}m(A) = \begin{cases} \dfrac{m(A)}{1 - {}^{\alpha}m(\Omega)} & \forall A \subset \Omega \\ 0 & \text{otherwise} \end{cases}, \tag{9} $$

The scenarios associated with this idea can often be found in the multi-evidence pooling systems, where decisions are made based on a set of existing pieces of evidence and the corresponding confidence in (or evaluation of) these pieces of evidence. Evidence and corresponding confidence may be elicited in different manners, e.g., drawn by different experts, or based on different viewpoints.

4 The suggestion of the model of e-shop selection processes the presence of price comparison sites

Suppose there are some categories of products (e.g., robotic vacuum cleaners that are manufactured by different companies and may have different functions). Let's denote the set of users $U = \{U_1, U_2, ..., U_n\}$, we suppose that these users have some referrals from friends, from recommendation systems from various information sources, or they prefer different brands of seeking item (robotic vacuum cleaner).

We further expect that this product is sold in various e-shops (the number of e-shops is m) at a price p_j. The price of delivery may also be important for users. It may vary depending on the price of the product. We include the price of delivery to the price of a product. Each e-shop is rated by customers after performed previous transactions with a certain number of points. This rating is available for other users to decide whether they will buy products

in certain e-shop or not. We can express these rating (preferences of e-shops) using a matrix $R=(r_{ij})_{n\times m}$ where r_{ij} represents user' (*j*) preferences (expressed by a certain number of points, for example in range 1 - 5) given to this e-shop (*i*).
The total number of point each for each of *m* e-shops is:

$$d_i = \sum_{j=1}^{n} r_{ji} \tag{10}$$

We have two parameters influencing the shopping behavior: the price of product and the reputation of e-shop expressed by the number of points obtained by various customers after previous transaction. We will combine these two parameters with the help of Dempster-Shafer theory [14].
We will express belief functions for the price parameter and for the reputation of e-shop.

Price parameter
This parameter shows how the shopping behavior of customers is influenced by the price. The respective belief functions have the following form:

$$\begin{aligned} m_p(\{buy_i\}) &= \alpha \frac{p_{\min}}{p_i} \\ m_p(\{\neg buy_i\}) &= 0 \\ m_p(\Theta_i) &= 1-\alpha \frac{p_{\min}}{p_i} \end{aligned} \tag{11}$$

where α is the weight of this evidence. We can intuitively read this weight as a reliability of this evidence, p_i is the price in the *i-th* e-shop, $p_{\min}$ is the minimal price at which the product is sold in respective e-shops. With this equation, we have expressed the tendency to purchase the product on the basis of the price. Usually, the lower the price the higher the temptation to buy the product in the *i-th* e-shop. At the same time, we assume that the equation is reflecting the temptation to buy the product, it does not show the reluctance to buy goods from the *i-th* e-shop, i.e. $m_p(\{\neg buy_i\}) = 0$.

Reputation parameter
This parameter shows how is the affection of customers to buy goods in the *i-th* e-shop is influenced by the reputation of respective e-shop. The belief functions have the following form:

$$\begin{aligned} m_R(\{buy_i\}) &= \beta \frac{d_i}{d_{\max}} \\ m_R(\{\neg buy_i\}) &= 0 \\ m_R(\Theta_i) &= 1-\beta \frac{d_i}{d_{\max}} \end{aligned} \tag{12}$$

where d_i is the total number of point each for each of *m* e-shops, $d_{\max}$ is the maximum pints assigned to some e-shop and where β is the weight of this evidence. With this equation, we have expressed the tendency to buy the product in a particular e-shop based on the good reputation of this e-shop. Usually, the better the reputation of the e-shop expressed by the number of points assigned to this e-shop by customers according their satisfaction with previous transactions. At the same time, we assume that the equation is reflecting the tendency to buy the product in certain e-shop, it does not show the reluctance to buy goods from the *i-th* e-shop, i.e. $m_R(\{\neg buy_i\}) = 0$.

The selection of e-shops without the influence of social network
Once we obtain the belief functions, we combine them in a consistent manner to get a more complete assessment of what the whole group of signs indicates. The combination of belief functions is done with the help of the Dempster's combination rule, see equation (4). We express the assumption that customer will buy certain product chosen by him/her in previous stage belief functions $m(\{buy_i\})$ in some e-shop *i*. We calculate the value $m(\{buy_i\})$ using the combination (see equation 4) of single belief functions expressing appropriate evidence for every of *i* e-shop:

$$m(\{buy_i\})=(m_p \oplus m_R)(\{buy_i\}) \tag{13}$$

We obtained the vector of $m(\{buy_i\})$ values as recommendation for user *U* to by certain product in *m* e-shops.

The selection of e-shops with the influence of social network

We use diffusion-based method called Network-Based Inference(NBI) [16] for the description of the influence of social network on decision making of some customer when he/she chooses an certain e-shop. For any customer u_i, all his e-shops *j* under his consideration o_j ($1 \le j \le n$) are sorted in the descending order. Then, we have defined the reinforcement (discount) coefficient γ_i based on the approach in [16] as follows:

$$\gamma_j = \sum_{l=1}^{n} w_{jl} a_{li} \tag{14}$$

We can the calculate the resulting belief about the choice of certain e-shop *i* $m(\text{buy}_j)$ by using the equation:

$$m_{R(\{buy_j\})} = \frac{m(\{buy_j\})}{1-\gamma_j} \tag{15}$$

Now, we create for any customer u_i, all his e-shops *j* under his consideration o_j ($1 \leq j \leq n$) sorted in the new descending order. Those e-shops with highest value of final resource are chose also on the basis of interaction with members of his social network.

Evaluation

In order to evaluate the accuracy and efficiency of recommendation algorithm we defined an evaluation criterion as follows:

$$AC = \frac{1}{n}\sum_{i=1}^{n}\frac{N}{L}, \tag{16}$$

where *n* is the number of customers, *L* is the length of the list of e-shops taking into consideration in experiments (less than the total number of e-shops available) and *N* is the number of correctly determined e-shops in the list of e-shops.

5 Research results

5.1 Data set description

In this paper, we used data obtained by experiments with the group of 58 students. They were told to buy certain product with the use of price comparison site Heureka.cz for this purchase. They had to proceed like in real situation, to consider price and reputation of e-shops. Then, they had to written a list of e-shops in the order in which they would wanted to buy certain product from these e-shops. They had a task to estimate the influence of peer in social network when choosing some e-shop.

On Heureka.cz, users grade e-shops using integers from one to five, which one stands for "extremely don't like" and five stands for "extremely like." We randomly divided the data set into two parts: 80% is the training set, and 20% is the testing set. To ensure the experimental results are accurate, all experiments were carried out with a fivefold cross validation test.

5.2 Experiment results

Taking reality into consideration, we could also assume the length L of the list of e-shops to be 5. We used the training set and the testing set to measure α and β by varying them from 0.0 to 1.0 step by 0.1 to see the values of *AC*. The aim was to determine values of α and β.

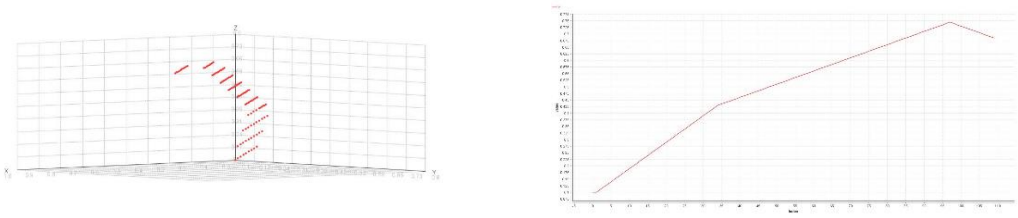


Figure 2 Value of *AC* criterion by varied α and β (Source: Own processing)

We have determined the best values α and β as following: $\alpha = 0.98$, $\beta = 0.89$ (see Figure1). Gamma values were quite different for individual students and e-shops in the range of about 0.65 to 1.39. However, the accuracy of the forecast was fairly good, around 90%.

6 Conclusions

We have suggested a model of e-commerce processes in the presence of a price comparison site. Experimental results on the group of students show accuracies of the proposed method. The accuracy of our model is sensitive to both amount of data and the length of list of e-shops. Our model can be useful at prediction of shopping behavior of customers in the presence of a price comparison site. In this paper, we focused mainly on the last part of the modeling process, where we used the belief function theory. However, it is clear, that the graph approach brings advantages over other methods as it also allows the expression of relationships between individual actors in the

field of e-commerce. Therefore, we want to focus on the use of network methods in the presented model in our next work.
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] Bellogin, A., Cantador, I., Castells, P.(2013) A comparative study of heterogeneous item recommendations in social systems, Inf. Sci. 221,142–169 .
[2] Beranek, L., & Nydl, V. (2013). 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.
[3] Beranek, L., Tlusty, P., & Remes, R. (2010). 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, Ceske Budejovice, Czech Republic, September 2010, s. 49-54.
[4] Beranek, L., Knizek, J. (2012). The Usage of Contextual Discounting and Opposition in Determining the Trustfulness of Users in Online Auctions. Journal of Theoretical and Applied Electronic Commerce Research, 7(1), 34 – 50
[5] Borgatti, S. P., & Halgin, D. S. (2011). Analyzing a affiliation networks. In The Sage Handbook of Social Network Analysis (Carring-ton, P. and Scott, J., eds.). The Sage handbook of social network analysis Thousand Oaks: Sage Publications, 417 - 433.
[6] Brynjolfsson, E. & Smith, M.D. (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers. Management Science 46, 563–585.
[7] Daniel, M. (2013) Properties of Plausibility Conflict of Belief Functions. In Rutkowski, L; Korytkowski, M; Scherer, R; et al., 12th International Conference on Artificial Intelligence and Soft Computing, ICAISC 2013, Zakopane, POLAND, Jun 09-13, 2013, PIEEE Computat Intelligence Soc, Poland Chapter, Lecture Notes in Artificial Intelligence, Vol. 7894, 235-246.
[8] Degeratu, A.M., Rangaswamy, A., & Wu, J. (2001). 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(1), 55–78.
[9] Easley, D., & Kleinberg, J. (2010). Networks, Crowds, and Markets. Cambridge: Cambridge University Press.
[10] Lamanauskas, V., Slekiene, V., Balog, A., & Pribeanu, C. (2013) Exploring the Usefulness of Social Networking Websites: a Multidi-mensional Model. Studies in Informatics and Control 22(2), 175-184.
[11] Pathak, B. K. (2012) Comparison Shopping Agents and Online Price Dispersion: A Search Cost based Explanation. Journal of Theo-retical and Applied Electronic Commerce Research. 7(1), 64 - 76.
[12] Rauch, J. E. (2010). 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.
[13] Schweitzer, F., Fagiolo, G., & Sornette, D. (2009). Economic Networks: The New Challenges. Science 325, 422.
[14] Shafer, G. (1976). A Mathematical Theory of Evidence. Princeton: Princeton University Press.
[15] Takayuki, M. & Tsutomu, W. (2013). Why Are Product Prices in Online Markets Not Converging? PLOS ONE. 8(8), 1 - 7.
[16] Zhou, T., Ren, J., Medo, M., Zhang, Y. (2007). Bipartite network projection and personal recommendation, Phys. Rev. E. 76 (4) 046115.

Age structure effects and consumption in the V4 countries

Terézia Bobošová[1], Luboš Střelec[2]

In this paper, the effects of a changing age distribution on aggregate consumption are analysed empirically. Growth in the number of older persons is a global phenomenon and do not omit countries of the Visegrad Group (V4). The analysis is done by estimating consumption functions of V4 countries which controls for age structure effects. The Life-Cycle Hypothesis posits that the main motivation for borrowing and saving is to smooth the consumption during different phases of people's life. Modigliani and his collaborators claim that individuals' consumption and saving behaviour are functions of their age; an individual borrows as young, saves as middle-aged and dissaves when old. Cross-country regressions regularly find that propensity to consume is higher when the population share of the elderly and children is high. The 'dependency ratio', defined as the number of children and retired persons to those of working age, is used to represent changes in the age structure, time series are annual. Time series for the economic variables consumption, income and wealth are quarterly and they cover the period from 2002(1) to 2017(4).

Keywords: Age structure, consumption, time series, Visegrad Group (V4) countries.

JEL Classification: C01, C32, E12, E21
AMS Classification: 91B15, 62P20, 62M10, 91B42

1 Introduction

Private consumption represents an essential part of aggregate demand and accounts for about two-thirds of the nation's Gross Domestic Product (OECD [22]). Not only due to this fact private consumption has been pleased with a substantive interest of economists, businessmen and politicians. The consumption of households includes both economic and social dimension. From the economic point of view at the microeconomic level, consumption encompasses all goods and services that meet the needs of consumers limited by their incomes. At macroeconomic level, consumption defines an important component of GDP as mentioned above. Consumption examining is also indirectly beneficial in the area of household savings analysis, as household income can be divided between consumption and savings. From a social point of view, private consumption is crucial for assessing living standards in conjunction with information on wealth and income (Stiglitz et al. [26]). The relationship between disposable income, consumption and savings is important and affects the overall economic climate, therefore there is a long tradition of studying the connection. The list of the variables that influence consumption together with the direction and magnitude of their effects presents the consumption function. Throughout history a large number of alternative hypotheses have been formulated.

Modigliani and Friedman introduced forward looking assumption in consumption function. Modigliani and Brumberg [21] proposed a Life-cycle Consumption Hypothesis (LCH) and stated that consumers save to smooth consumption over a lifetime. Further, Friedman's [13] Permanent Income Hypothesis (PIH) says consumption only depends upon the shock in the permanent part of income but not in transitory income.

LCH predicts that the age composition of the population affects consumption and savings. The aim of this this paper is to investigate empirically whether changes in the age structure of the population affect aggregate consumption of countries belonging to the Visegrad Group. The econometric model of the consumption function which takes the age structure effects into account is developed.

The paper is organized as follows: section 2 brings a brief overview on the theoretical approach to consumption function intent on Life-cycle Consumption Hypothesis and the recent literature discussion in the context of age structure. Section 3 introduces data and methodology. The results of the empirical analysis are presented in section 4. Section 5 is dedicated to conclusions of the analysis.

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Economics, Zemědělská 1, 613 00 Brno, xbobosov@node.mendelu.cz.
[2] Mendel University in Brno, Faculty of Business and Economics, Department of Economics, Zemědělská 1, 613 00 Brno, lubos.strelec@mendelu.cz.

2 Literature overview

Consumer spending is the largest single expenditure category in the final demand for goods and services, accounting for about two-thirds of gross domestic product (GDP). For the reason households' choice of how much of their income to consume and save is one of the most discussed literature topics in macroeconomics. A clear understanding of the underpinnings of consumer spending is a valuable asset for central bankers and policymakers (Chatterjee [6]). Along the years a large number of alternative hypotheses have been formulated and modelling of consumption function can be considered a constantly evolving process.

One of the most influential theories by far is The Life-Cycle Hypothesis (LCH), first developed by Modigliani and Brumberg [21]) and later extended by Ando and Modigliani [1]. LCH is based on the assumption that households plan their consumption at the horizon of their life-span. The early empirical tests of the LCH were tests of whether wealth and the interest rate explained consumption better than current disposable income. Consumer's spending is funded by the lifelong working income and the initial wealth. The working income is in terms of this theory changing in a predictable way, depending on the different stages of the life-cycle (Parker [23]). The main limitations of the consumers are the length of their lives and the resources available over their lives. The basic factor affecting the relationship between consumption expenditure and income is the age of an individual – people make intelligent choices about how much of their income they want to spend at each age. In the young adulthood and retirement phases, when the received income is low, consumption patterns are maintained through recourse to borrowing or by drawing down past savings. Consequently, in these phases of the life cycle, consumption is a high proportion of income. As regards the middle phase, when income tends to be relatively high, savings are built up to finance post-retirement consumption with the result that a smaller proportion of income is consumed in this phase (Deaton [7]), (Bayar and Mc Morrow [3]).

The relationship between saving and the age-structure of the population is also a current topic of debate. Cross-country regressions regularly find that aggregate saving rates are lower when the population share of the elderly and the population share of children are high. Such regressions have strong negative predictions for currently high saving countries whose populations are aging, particularly countries in South and East Asia (Deaton [7]).

The aggregate consumption function of the LCH is not very different in practice from those developed from other approaches, particularly Milton Friedman's PIH [13]. Changes in current income influence changes in current consumption only to the extent that such changes can be regarded as being permanent and consequently justify a recalculation of lifetime consumable resources. In the event of temporary income gains, the consumption impact is likely to be small. Both hypotheses are seen as complementary. PIH for the main emphasize the way individuals shape expectations of their income and LCH focuses primarily on the role of demographic variables in consumer decision-making.

The share of older persons in the total population will increase significantly in the coming decades, as a greater proportion of the post-war baby-boom generation reaches retirement. Nowadays the numbers of elderly are increasing in European countries and the share of older persons in the total population will even increase significantly in the coming decades. The proportion of people of working age is shrinking while the relative number of those retired is expanding (Eurostat [11]). Will the ageing of V4 countries bring down its saving rate and bring its growth to a halt? The link between the propensity to consume or savings rate and age distribution with focus on the issue whether the elderly dissave, has been widely explored in the empirical literature.

Economic theory suggests that the changes in the age structure of the population affect macroeconomic variables such as aggregate consumption and the savings rate. Studies based on aggregate macroeconomic data usually confirm the predictions of the LCH. Many recent studies have been motivated by the observed reduction in consumption following retirement, spotted in many industrialised economies (Banks et al. [2]), (Bernheim et al. [5]).

Heien [15] Masson et al. [20], Horioka [17], Higgins [16], Gourinchas [14], Senesi [25], Erlandsen and Nymoen [10] and Kuhn and Prettner [19] have found out that share of savings decrease or share of aggregate consumption rises when the share of elderly persons in the population increases. Fair and Dominguez [12] have specifically found that prime-age people consume less relative to their income than other age groups on US data what is in line with the predictions of the LCH. Heien [15], who has estimated a multiperiod consumption function on annual U.S. data, has also reported a significant relationship between the age structure of the population and consumption of non-durables and services. Berg [4] has investigated the relationship between the population age distribution and consumption on Swedish annual data. He has concluded that different age shares of the population have significant effects on the demand because of private consumption and reported that the estimated coefficients of the different age shares variables are consistent with the predictions of the LCH.

Not all studies found significant age composition effects. Denton and Spencer [8] have not supported the idea that age distribution of the population influence consumption significantly. Eguia and Echevarria [9] have studied the effect of population age distribution upon private consumption expenditure in Spain following the method in Fair and Dominguez [12]. They have discovered that the pattern of the coefficients of the demographic variables is not consistent with the LCH: instead of showing a stable level of consumption throughout the life cycle, younger and elderly individuals consume the least; middle aged individuals spend the most. The research in Serbia (Radivojević and Vasić [24]) has also shown that household head age is a significant factor in the formation of the total amount and structure of household consumption. The age of the household head has a direct influence on household income level and determines affects on the level and structure of household consumption.

3 Data and methodology

The research is based on the regression analysis using quarterly data of Visegrad Group countries covering the period from 2002(1) to 2017(4). In national accounts, final consumption expenditure of households denotes expenditure on goods and services that are purchased and paid for by households. It is an indicator describing the material welfare of households. In order to estimate real private consumption, following explanatory variables are used:

- gross domestic product (W),
- compensation of employees (Y),
- dependency ratio (AGE),
- real interest rates (IR).

Real GDP denotes a proxy of wealth. GDP is the central measure of national accounts, which summarises the economic position of a country. GDP using chain-linked volume in million euro (real terms, price level of 2010) is intended to allow comparisons of the economic development both over time and between V4 economies. Wealth is considered to be one of the most important determinants of consumption expenditures from the perspective of the LCH. If the LCH is approved, then real GDP as a proxy of wealth should manifest a positive sign.

Compensation of employees signs a proxy of current income. It is defined as the total remuneration, in cash or in kind, payable by an employer to an employee in return for work done by the latter during the accounting period. Compensation of employees consists of wages and salaries and of employers' social contributions. The data are expressed using quarterly figures chain linked volume in million euro (price level of 2010). If current income is significant while lagged insignificant, it is in accordance with Keynes' Absolute Income Hypothesis [18], not LCH.

The age structure variable the 'dependency ratio', defined as the number of children (population aged 0-19) and retired persons (population aged 60 and more) to those of working age (population aged 20–59), is used to represent changes in the age structure. Age structure variable is included as follows:

$$AGE_t = \left(\frac{\text{population aged } 0-19 + \text{population aged 60 and more}}{\text{population aged 20–59}}\right)$$

Given the prediction that young and old persons save less (consume more) relative to their income than middle aged population, AGE is expected to enter the consumption function with a positive coefficient.

Another variable that may influence consumption in the long run is the real interest rate. Real interest rates are connected with the possibilities deriving from financial market as lending and borrowing. However, there may also be substitution effects from interest rate changes; an increase in real interest rates makes consumption today more expensive relative to tomorrow's consumption; hence, consumption is expected to decline, savings are more attractive. However, higher interest rates increase financial wealth and even a positively signed coefficient can result.

The purpose of the empirical model is to look at the effects of population ageing on household consumption, including other factors that might also affect the consumption expenditures of households of V4 countries. The regression model based on the abovementioned theoretical and empirical studies of consumption function is specified by following relationship:

$$C_t = \mathrm{f}(W_t, Y_t, AGE_t, IR_t)$$

For seasonal adjustment the TRAMO analysis is used. TRAMO/SEATS is recommended by ESS Guidelines on Seasonal Adjustment and officially used by Eurostat, from where this dataset also come.

4 Results

Preliminary to the estimation process, all variables have been tested for unit roots using augmented Dickey-Fuller (ADF) test. All variables have been found to be non-stationary in level because of the seasonal character. The relationship between non-stationary time series is associated with the problem of spurious regression, when regression results become unreliable and the resulting models generally have poor forecasting ability. The only exception is if the model eliminates the stochastic trends to produce stationary residuals cointegration.

Czech Republic:

$$C_t = -180.32 + 0.792\mathrm{C}_{t-1} + 0.145\mathrm{W}_t - 0.076W_{t-1} + 33.067AGE_t - 167.634D_t$$

D=1 for t=40, 41, …, 63 (D=1 for period: 2012:1–2017:4)

Hungary:

$$C_t = 1650.59 + 0.628\mathrm{C}_{t-1} + 0.344\mathrm{Y}_t - 48.32IR_t + 515.748D_t - 665.075Q_t - 252.216Z_t$$

D=1 for t=9, 10, …, 63 (D=1 for period: 2004:2–2017:4)

Q=1 for t=10, 11, …, 63 (Q=1 for period: 2004:3–2017:4)

Z=1 for t=26, 27, …, 63 (Z=1 for period: 2008:3–2017:4)

Poland:

$$C_t = -1840.35 + 0.733\mathrm{C}_{t-1} + 0.043\mathrm{W}_t + 0.119W_{t-1} + 54.374AGE_t - 833.517D_t$$

D=1 for t=40, 41, …, 63 (D=1 for period: 2011:4–2017:4)

Slovakia:

$$C_t = -3483.57 + 1.111\mathrm{C}_{t-1} + 0.294\mathrm{W}_t - 0.163W_{t-1} - 0.423Y_t + 74.156AGE_t$$

From the Table 1 is evident, that the residues of every model are stationary, the time series are co-integrated, which excludes the possibility of spurious regression. However, since dummy variables for sudden shifts in consumption expenditures enter the system, residuals behave as a white noise sequence.

Variable	Levels Statistics	p-value
Residual of the model of Czech Republic	-7.067	0.000
Residual of the model of Hungary	-9.236	0.000
Residual of the model of Poland	-4.772	0.000
Residual of the model of Slovakia	-2.731	0.006

Table 1 ADF test of residues of models

The empirical relationship among aggregate consumption and disposable income, wealth, age structure and interest rates has shown that changes in the age distribution of the population have significant and life-cycle-consistent effects on aggregate consumption. The parameter of propriate variable AGE, expressed as the dependency ratio, met the expectations and influenced the consumption function with a positive coefficient. The only exception is Hungary, where results showed a dependence of real private consumption on current income without statistical effects of age distribution of population, which showed support for the Keynes' Absolute Income Hypothesis [18] in contrast to other V4 countries.

5 Conclusions

In this paper, the empirical relationship between aggregate consumption and the age structure of the population in V4 countries have been investigated. The analysis is based on aggregate time series data and age structure changes has been represented by the dependency ratio. For the purpose of the private consumption function estimates dynamic regression models has been used. It has been shown that changes in the age structure of the population affect aggregate consumption significantly, and that the age structure impact is in line with economic theory. More specifically, the results support the predictions of the Life-Cycle Hypothesis, which imply that young adults and old persons have a higher average propensity to consume than the middle-aged. Although several studies have reported significant age structure effects on aggregate consumption in different countries.

Changing age distribution can play a key role in the consumption function. Since the share of older persons in the population of V4 countries is blowing up, the differences of the forecasts from this model might be actual. In contrast to economic variables, demographic variables are quite easy to forecast in the short and medium run.

Acknowledgements

Supported by the grant No. PEF_DP_2018030 of the Grant Agency IGA PEF MENDELU.

References

[1] Ando, A. & Modigliani, F. (1963). The "Life cycle" hypothesis of saving: aggregate implications and tests. *Economic Review*, **53**, 55–84.
[2] Banks, J. & Blundell, R. & Tanner, S. (1998). Is there a retirement-savings puzzle? *American Economic Review*, **88**, 4, 769–88.
[3] Bayar, A. & McMorrow, K. (1999). Determinants of private consumption, EU Commission - Working Document. *Economic Papers*, **135**.
[4] Berg, L. (1996). Age distribution, saving and consumption in Sweden. *Working paper*, **22**, Uppsala University.
[5] Bernheim, B. & Skinner, J. & Weinberg, S. (1997). What accounts for the variation in retirement wealth among US households? *NBER Working Paper*, National Bureau of Economic Research, Cambridge, Mass.
[6] Chatterjee, S. (2009). The Peopling of Macroeconomics: Microeconomics of Aggregate Consumer Expenditures. *Federal Reserve Bank of Philadelphia*.
[7] Deaton, A. (2005). Franco Modigliani and the Life Cycle Theory of Consumption. *Banca Nazionale del Lavoro Quarterly Review*, **58**.
[8] Denton, F. T. & Spencer, B. G. (1976). Household and population Effects on Aggregate Consumption. *Review of Economics and Statistics*, **58**, 1, 86–95.
[9] Eguia, B. & Echevarria, C. A. (2001). Population age structure and housing investment in Spain. Mimeo.
[10] Erlandsen, S. & Nymoen, R. (2008). Consumption and population age structure. *Journal of Population Economics*, **21**, 505–520.
[11] Eurostat. (2017). Population structure and ageing: tables and figures.
[12] Fair, R. C. & Dominguez, K. (1991). Effects of the changing U.S. age distribution on macroeconomic equations. *American Economic review*, **81**, 5, 1276–1294.
[13] Friedman, M. (1957). *A theory of Consumption Function*. 1. ed. Princeton University Press. ISBN 0-691-04182-2.
[14] Gourinchas, P-O., & Parker, J. (2002). Consumption Over the Life Cycle. *Econometrica*, **70**, 47–89.
[15] Heien, D. M. (1972). Demographic Effects and the Multiperiod Consumption Function, *Journal of Political Economy*, **80**, 1, 125–138.
[16] Higgins, M. (1998). Demography. National savings and international capital flows. *International Economic Review*, **39**, 2, 343–369.
[17] Horioka, C, Y. & Okui, M. (1999). A U.S.-Japan comparison of the importance and determinants of retirement saving. *Economics Letters*, **65**, 3, 365–371.
[18] Keynes, J.M. (1936). *The general theory of employment, interest, and money*. United Kingdom: Palgrave Macmillan. ISBN 978-0-230-00476-4.
[19] Kuhn, M. & Prettner, K. (2018). Population age structure and consumption growth: evidence from National Transfer Accounts. *Journal of Population Economics*, **31**, 135–153.
[20] Masson, P. R. & Bayoumi, T. S. & Samiei, H. (1998). International evidence in the determinants of private saving. *The World Bank Economic Review*, **12**, 3, 483–501.
[21] Modigliani, F. & Brumberg, R. (1954). Utility analysis and the consumption function: an interpretation of the cross-section data, in K. Kurihara (ed.). *Post-Keynesian Economics*. Rutgers University Press. New Brunswick, NJ, 388–436.
[22] OECD. (2010). National Accounts at a Glance 2009. Paris. Available online from the OECD Publishing: http://www.oecd-ilibrary.org/content/book/9789264067981-en
[23] Parker, J. (2010). 16 Theories of Consumption and Saving. *Economics 314 Coursebook*.
[24] Radivojević, B. & Vasić, P. (2012). Household age structure and consumption in Serbia. *Economic Annals*, **57**, 195, 79–102.
[25] Senesi, P. (2003). Population Dynamics and Life-Cycle Consumption. Journal of Population Economics, **16**, 389–394.
[26] Stiglitz, J. & Sen, A. & Fittoussi, J-P. (2009). Commission on the Measurement of Economic Performance and Social Progress. Paris: *Performance and Social Progress*.

Effect of risk measure on the portfolio composition made by fuzzy return-risk model: Real case with open unit trusts

Adam Borovička[1]

Abstract. Investment decision making can be difficult because is usually influenced by many factors. It is no doubt that one of the most crucial factors is a risk of the investment. The risk can be measured in several ways. Many of them are based on the historical development of the asset's return (variance, semivariance, average absolute negative deviation, etc.). The existence of many concepts raises the essential question whether a type of risk measure could influence an investment decision, the composition of an investment portfolio (and its characteristics) respectively. Answer this question can have a significant impact on the whole investment decision making process. Therefore, this problem is analyzed in this paper. For making a portfolio, a fuzzy return-risk model is applied to take into account a typically unstable return of assets. To make a complex and representative analysis, the effect of risk measure is performed for various investment time period and several groups of open unit trusts that are subject of this empirical study. The resulting portfolios are compared to clearly declare the impact of chosen risk measures on the portfolio composition.

Keywords: effect of risk measure, fuzzy risk-return model, open unit trust, portfolio composition, (triangular) fuzzy number.

JEL Classification: C44, G11
AMS Classification: 90B50

1 Introduction

Investing is more and more popular in the Czech Republic. The investment in open unit trusts is not an exception. The investment decision is usually influenced by various criteria. It is no doubt that one of the most important criterion is a risk of the investment. Risk is usually comprehended as a threat of investment loss. It can be measured in many ways. A risk measure usually depends on a historical development of the asset's return. The well-known concepts are standard deviation, variance, semivariance, average absolute negative deviation, variation coefficient, beta coefficient, etc. Then, one of the important decisions in terms of the investment decision making process (investment portfolio selection) is the choice of an appropriate technique for a risk measurement. Not only from my point of view, the most important criterion of a selection of risk measurement technique is its intelligibility and interpretability of its value, as well as its subsequent (technical) usability for a portfolio selection. Through this decision making, a brighter (potential) investor could wonder whether this choice could influence the measurement of the risk, and hence a composition of the portfolio. If the effect is insignificant (the portfolios are the same or very similar), the most important remains an aforementioned user-friendliness of the risk measure concept. Otherwise, the effect should be considered, because a different portfolio composition provides a different level of monitored portfolio characteristics (e.g. return). Especially, significant impact on the portfolio return can essentially influence a selection of the used risk measure (and consequently portfolio selection). Therefore, the effect of risk measure on the investment portfolio composition is complexly analyzed in this article.

To select a portfolio, an optimization approach inspired by modern Markowitz portfolio theory is used [8,9]. The Markowitz mean-variance optimization model is very convenient supporting tool for a portfolio selection. At first, the optimization goal can be easily changed (risk or return). Secondly, a basic model can be extended by other conditions (e.g. for "quantitative" diversification, expression of the additional investor's preferences). Further, a standard mean-variance concept can be transformed to more general 2-factors *return-risk* form. This model can take into account a typical element of uncertainty – unstable return. Then the return is seemly quantified by a (triangular) fuzzy number (e.g. in [3] or [10]). Such a flexible fuzzy return-risk model is applied to satisfactorily make a risk effect analysis.

As indicated above, for a risk effect analysis, the open unit trusts (bond and stock), managed by Erste-Sparinvest and offered through Česká spořitelna, are selected. From historical observations, the instable return of unit trusts is formulated as a triangular fuzzy number. Risk is measured by the most commonly used tools as standard deviation, variance, semivariance, average absolute negative deviation and SRRI. The main aim of this paper is to

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

complexly analyze the effect (impact) of a risk measure on the composition of a portfolio (with open unit trust in this case). For this purpose, a methodological approach based on a fuzzy risk-return model is proposed. Then the effect is studied in various situations determined by a kind of the open unit trusts or investment time period. For all situations, the mathematical programming models are formulated and solved. The resulting portfolios are compared from the point of view of their composition and particular characteristics. The results of such a complex yet unplanned empirical study with these funds can help to choose a "right" risk measure, or make a portfolio selection process more effective.

The structure of the paper is as follows. After the introduction (Section 1), a methodological approach for an analysis of risk effect is introduced in terms of Section 2. Fuzzy return-risk model is described, the selected risk measures as well. In Section 3, the mentioned above analyses with open unit trusts are carried out. The results are discussed. Conclusion (Section 4) summarizes the article and provides some ideas for future research.

2 Methodological approach for an analysis of risk measure effect

In this section, the entire methodological approach for an analysis of risk measure effect is described. In the first phase, risk measures for the analysis are selected. In terms of the second phase, the mathematical models (fuzzy return-risk models) for portfolio selection in each specified situation (by investment time period, or type of investment instrument) are formulated and subsequently solved. Finally, to trace the effect of risk measure type on a portfolio composition, solutions of all investment cases are compared and analyzed.

2.1 Risk measure concepts

The most often risk measures (in investment) are logically selected for this analysis (variance, or standard deviation, semivariance). From my point of view, better and probably not so well known and often used average absolute negative deviation concept is also introduced. Finally, the official point risk measure published by banks should not be omitted. It is called Synthetic Risk and Reward Indicator (SRRI).

Variance/standard deviation, semivariance

One of the most often used concept of risk measurement is a statistic characteristic *variance* (or *standard deviation*). The variance, or standard deviation of return of the *j*-th open unit trust is calculated as follows [11]

$$\sigma_j^2 = \frac{\sum_{i=1}^{m}(x_{ij} - \overline{x}_j)^2}{m}, \text{ or } \quad \sigma_j = \sqrt{\frac{\sum_{i=1}^{m}(x_{ij} - \overline{x}_j)^2}{m}} \qquad j = 1,2,...,n, \tag{1}$$

where $\overline{x}_j$, $j = 1,2,...,n$, denotes an average return, $x_{ij}, i = 1,2,...,m;\ j = 1,2,...,n$ is the *i*-th return, *m* is the number of historical returns of the *j*-th open unit trust. Risk of the portfolio r_p is calculated as variance, or standard deviation, in the following form [1]

$$r_p = \sigma_p^2 = \sum_{k=1}^{n}\sum_{l=1}^{n} x_k x_l \sigma_{kl}, \text{ or } \quad r_p = \sigma_p = \sqrt{\sum_{k=1}^{n}\sum_{l=1}^{n} x_k x_l \sigma_{kl}}, \tag{2}$$

where $\sigma_{kl}, k, l = 1,2,...,n,$ is a covariance (variance) of returns between the *k*-th and *l*-th open unit trust, $x_k, k = 1,2,...,n,$, or $x_l, l = 1,2,...,n,$ denotes a share of the *k*-th, or *l*-th fund in the portfolio. This function is not linear. Therefore, finding the optimal solution (global extreme) of model below (7) may not be easy. If a portfolio is composed by one homogenous type of investment instrument, the covariances do not usually play a significant role. Then they can be eliminated, and simpler form of a portfolio risk measure can be formulated by means of the weighted sum as in the model (6)

$$r_p = \sum_{j=1}^{n} \sigma_j^2 x_j, \text{ or } \quad r_p = \sum_{j=1}^{n} \sigma_j x_j, \tag{3}$$

where x_j, $j = 1,2,...,n$, represents a share of the *j*-th open unit trust in a portfolio. The variance (or standard deviation) concept is burdened with a significant drawback. Under risk minimization, it is undesirable to include positive deviations from the average return. This disadvantage can be weakened by *semivariance* concept [9]

$$sem\sigma_j^2 = \frac{\sum_{x_{ij} < \overline{x}_j} (x_{ij} - \overline{x}_j)^2}{m}, \tag{4}$$

where *m* is the number of historical returns less than an average return. Risk of the portfolio can be expressed in a linear shape $r_p = \sum_{j=1}^{n} sem\sigma_j^2 x_j$.

Average absolute negative deviation

From my point of view, the square roots in a variance concept are unnecessary. Then risk can be measured as average absolute negative deviation inspired by Konno and Wijayanayake [7]

$$AAND_j = \frac{\sum_{x_{ij} < \bar{x}_j} (\bar{x}_j - x_{ij})}{m}, \tag{5}$$

where *m* is the number of historical returns less than average return. I think that this concept is the most comprehensible for users. Its calculation and interpretation are easy. Risk is presented as average negative deviation from the mean. A portfolio risk is specified as the weighted sum is in the previous cases.

Synthetic Risk and Reward Indicator (SRRI)

This measure is proposed by Committee of European Securities Regulators [5]. This indicator takes the values on a scale from 1 to 7 (1-the lowest risk, 7-the highest risk). Each grade corresponds with a particular scale of risk measured as a standard deviation of the returns covering the last 5 years of the fund's life. Risk of the portfolio is standardly measured as a weighted sum $r_p = \sum_{j=1}^{n} SRRI_j x_j$, where $SRRI_j, j = 1,2,...,n,$ is the value of SRRI of the *j*-th open unit trust.

2.2 Fuzzy return-risk optimization model

The mean-variance model for a portfolio selection is proposed by Markowitz [8]. This approach has a few following advantages. This tool is user-friendly. The mathematical model is simple. Its optimal solution can be easily found. Two most important characteristics are included, i.e. risk and return. Moreover, this model can be extended by other necessary conditions for a particular portfolio selection situation. Original mean-variance risk can be generalized to the form return-risk where namely risk is expressed as another indicator than variance (standard deviation, average absolute negative deviation, etc.). Two crucial factors are kept behind. However, an original diversification idea naturally takes away by using weighted sum approach for a risk measurement. Such a model is not able to consider an unstable (uncertain) return of the open unit trust as a typical element on the capital market. This element can be effectively quantified by means of a (triangular) fuzzy number. Then a general (extended) fuzzy *return-risk* model can be formulated as follows

$$\begin{array}{lll} z(x_1,...,x_n) = \sum_{j=1}^{n} \tilde{v}_j x_j \to \max & & z(x_1,...,x_n) = \left(\sum_{j=1}^{n} v_j^l x_j, \sum_{j=1}^{n} v_j^m x_j, \sum_{j=1}^{n} v_j^u x_j \right) \to \max \\ \sum_{j=1}^{n} r_j x_j \le R & \text{, or} & \sum_{j=1}^{n} r_j x_j \le R \\ \mathbf{x} \in X & & \mathbf{x} \in X \end{array} \quad , \tag{6}$$

where $\tilde{v}_j = (v_j^l, v_j^m, v_j^u), j = 1,2,...,n,$ is a return of the *j*-th open unit trust expressed as a triangular fuzzy number represented by 3 standard parameters the lowest v_j^l , middle (average) v_j^m and the highest v_j^u return (from historical sample) of the *j*-th fund. $r_j, j = 1,2,...,n,$ represents a risk and $x_j, j = 1,2,...,n,$ is a share of the *j*-th fund creating the vector $\mathbf{x} = (x_1,...,x_n)^T$. The set *X* is composed by constraints mathematically expressing the portfolio requirements (e.g. upper/lower limit of share of the open unit trust in the portfolio). Necessary constraint is a "portfolio condition" $\sum_{j=1}^{n} x_j = 1$. Return of the portfolio as a triangular fuzzy number can be clearly seen in the formulation on the right side. This model is formulated as maximizing, because the return is usually considered as the most important criterion. However, the importance of a risk can be expressed by its upper limit *R* that can be determined based on the knowledge of the lowest and highest possible risk. These extreme values can be stated by means of one-objective model specified for minimizing, or maximizing risk function $z_r(\mathbf{x})$ on the set *X*. Let $Z_{r_l}^* = Z_r(\mathbf{x}_\mathbf{l}^*)$, $Z_{r_u}^* = Z_r(\mathbf{x}_\mathbf{u}^*)$ denote the lowest, or the highest possible risk. Then the following holds

$$\mathbf{x}_\mathbf{l}^* = \arg\min_{\mathbf{x}\in X} z_r(\mathbf{x}), \qquad \mathbf{x}_\mathbf{h}^* = \arg\max_{\mathbf{x}\in X} z_r(\mathbf{x}). \tag{7}$$

For upper limit of risk level holds $R = z_{r_l}^* + p(z_{r_u}^* - z_{r_l}^*)$, where $p \in \langle 0,1 \rangle$ relatively represents a position between both extreme values. To solve the problem (6), the fuzzy goal principle [2] is applied through optimization technique maximin [12]. Thus, model (6) is transformed to the following strict form

$$\begin{aligned} & z = \alpha \to \max \\ & \frac{\sum_{j=1}^{n} v_j^l x_j - z_{min}^{l*}}{z_{max}^{l*} - z_{min}^{l*}} \geq \alpha \qquad \frac{\sum_{j=1}^{n} v_j^m x_j - z_{min}^{m*}}{z_{max}^{m*} - z_{min}^{m*}} \geq \alpha \qquad \frac{\sum_{j=1}^{n} v_j^u x_j - z_{min}^{l*}}{z_{max}^{u*} - z_{min}^{u*}} \geq \alpha, \\ & \sum_{j=1}^{n} r_j x_j \leq R \qquad 0 \leq \alpha \leq 1 \qquad \mathbf{x} \in X \end{aligned} \tag{8}$$

where $z_{min}^{l*}, z_{\min}^{m*}, z_{min}^{u*}$, or $z_{max}^{l*}, z_{max}^{m*}, z_{max}^{u*}$ are the lowest, or the highest possible return of a portfolio accordant with the value of particular parameter of triangular fuzzy numbers representing a vague returns of open unit trusts. These extreme values are determined by means of similar models as (7).

All formulated mathematical models are linear. The set of feasible solution is limited (at least thanks to a portfolio and non-negativity conditions). Then the optimal solution can be easily found by a simplex method.

2.3 Risk effect analysis

In the last stage, an empirical analysis of risk effect can be done. We have solutions of all needed models (7) and (8) representing situations with diverse groups of selected open unit trusts, various investment time periods and types of risk measures. The same solutions (in each category or globally) indicate independency of a portfolio composition on a risk measure type. If the solutions are different (i.e. different composition of portfolios), the effect of risk measure is noticeable. Then an interesting question should be whether some applied risk measure systematically allows to gain a better level of expected portfolio return (under some specified threshold value of risk) than the others. To compare the returns, a modified McCahone's approach can be applied [4]. A complexity of performed analysis allows to study a demonstration of various investment time periods and kind(s) of selected open unit trusts on a risk effect.

3 Analysis of effect of risk measure on a composition of the portfolio with open unit trusts

The investment in open unit trusts become popular in the Czech Republic in last two decades. Česká spořitelna is one of the biggest manager and "issuer" of the open unit trusts in our country. Accept own open unit trusts, this company also arranges trading with open unit trusts managed by Erste-Sparinvest. A selection of these open unit trusts is also supported by the author's personal experiences. For risk effect analysis, four bond and five stock open unit trusts (see in Table 1) are selected. Based on many empirical observations, the most important criteria are (expected) return and risk of the investment. Return is usually estimated based on the historical observations. Risk is also measured through the historical returns as some statistical characteristics mentioned above. This investment in open unit trusts is rather considered as longer-time. Therefore, the funds with a sufficient history are chosen. However, a complex analysis of the effect of risk measure should be carried out for other time (non-standard) period. Such an interesting period can be January (connected with *January effect*) and May-October period (connected with *Sell in May and go away effect*). In each period, the analysis is performed for both groups of funds, as well as for all funds together. In each investment case, all five risk measures are tested. Such a complex empirical analysis can provide representative results to judge the effect of risk measure on a portfolio composition.

3.1 Standard longer-time investment period

Historical period from January 2006 to March 2018 is selected. This period should represent longer development of the prices (returns) of the open unit trusts. It includes various situation on the capital market (huge falls, big rises, "calmer" period). All data about yearly returns and risk are displayed in the following table (Table 1). The parameters for fuzzy returns are calculated based on the prices of allotment certificates of open unit trusts collected from the charts on the Investment center web side [6], the SRRI data as well. Besides the SRRI scale 1-7, all data is in percentage.

Open unit trust	Return	Var.	Semivar.	St. dev.	AAND	SRRI
Portfolio Bond Europe	(−6.53, 2.30, 12.31)	34.03	35.18	5.83	4.89	3
Český fond státních dluhopisů	(−4.61, 2.42, 11.45)	18.28	15.65	4.28	3.40	3
Český fond firemních dluhopisů	(−2.61, 0.78, 5.36)	5.24	5.70	2.29	2.16	2
Bond Europe-High Yield	(−34.50, 6.05, 57.57)	405.45	275.20	20.14	11.29	3
Stock Japan	(−29.56, 0.86, 31.36)	282.11	367.20	16.80	16.38	6
Stock Istanbul	(−63.82, 8.22, 77.94)	1803.03	1755.75	42.46	38.34	7
Stock Global	(−53.48, 4.54, 33.76)	464.54	619.68	21.55	15.18	5
Stock Europe Property	(−42.26, 5.85, 41.99)	608.94	1122.85	24.68	31.07	6
Stock Europe Emerging	(−74.24, 3.68, 58.95)	1069.87	1286.44	32.71	25.71	6

Table 1 Data for a standard longer-time investment period

Based on (not only) author's experiences, the investors usually want to have a higher number of unit trusts in their portfolio to partly diversify it. Therefore, a maximum level of the share of one open unit trust is determined at the level of 30 %. The investment strategy is based on the maximizing portfolio return with demand on a maximum level of the portfolio risk. Because risk is also important investment criterion, this threshold is stated at rather lower level. Under these conditions, a portfolio of open unit trusts for a standard longer-time investment horizon is made through the following strict (defuzzified) model

$$\begin{aligned} & z = \alpha \to \max \\ & \frac{\sum_{j=1}^{9} v_j^l x_j - z_{min}^{l*}}{z_{max}^{l*} - z_{min}^{l*}} \geq \alpha \qquad \frac{\sum_{j=1}^{9} v_j^m x_j - z_{min}^{m*}}{z_{max}^{m*} - z_{min}^{m*}} \geq \alpha \qquad \frac{\sum_{j=1}^{9} v_j^u x_j - z_{min}^{l*}}{z_{max}^{u*} - z_{min}^{u*}} \geq \alpha \quad , \\ & \sum_{j=1}^{9} r_j x_j \leq z_{r_l}^* + 0.2(z_{r_h}^* - z_{r_l}^*) \qquad 0 \leq \alpha \leq 1 \qquad \mathbf{x} \in X \end{aligned} \tag{8}$$

where x_j $(i = 1, 2, ..., 9)$ is a share of the *j*-th open unit trust in the portfolio (the values of index *i* related to the funds in order from Table 1) and $X = \left\{ \mathbf{x} = (x_1, ..., x_9)^T \mid \sum_{j=1}^{9} x_j = 1; 0 \leq x_j \leq 0.3,\ j = 1, 2, ..., 9 \right\}$. Models (8), or (7) are solved for each risk measure for all nine open unit trusts, as well as for both group separately. For the case of all funds together, the portfolios are as follows (Table 2).

Open unit trust	Variance	Semivar.	St. dev.	AAND	SRRI
Portfolio Bond Europe	7.92 %	13.85 %	30 %	25.19 %	22 %
Český fond státních dluhopisů	30 %	30 %	30 %	30 %	x
Český fond firemních dluhopisů	x	x	6.42 %	x	30 %
Bond Europe-High Yield	30 %	30 %	30 %	30 %	30 %
Stock Japan	21.60 %	18.25 %	x	x	x
Stock Istanbul	x	7.90 %	3.58 %	x	18 %
Stock Global	x	x	x	x	x
Stock Europe Property	10.48 %	x	x	x	x
Stock Europe Emerging	x	x	x	14.81 %	x

Table 2 Portfolios for all 9 open unit trusts

The portfolios are diverse. Apart from one fund, they are all represented. A portfolio diversity is mainly caused by a type of risk measure. As we can see in Table 1, the "risk relationship" among the funds can be significantly different for risk measures (e.g. variance and semivariance for Stock Istanbul and Global Stock - for one higher variance, for another higher semivariance). As expected, risk of the portfolios is at the highest possible value because the investment with a higher risk usually provides a higher expected return. Then a comparison of the portfolios in terms of the return is meaningful. The best return has the portfolio made by help of SRRI risk indicator. This situation holds for another two groups of funds. However, SRRI indicator is calculated from 5-year data where there are not such turnover than in the last 12 years. Ranking in other places is diverse. Similar result is for stock open unit trusts. A slightly different situation is for bond funds where 3 portfolios are the same (for variance, semivariance and standard deviation). This fact is mainly caused by a similar historical price development. But

a situation with the portfolio returns is similar. To conclude, a risk measure significantly affects a portfolio composition. The most "suitable" risk measure (from the perspective of return) can be selected in some concrete investment situation. However, no applied risk measure ensures an achievement of the highest expected return at a more general level.

3.2 Special investment periods – *January* and *Sell in May and go away* effect

The *January effect* lies in the fact that the stock returns in this month are statistically above average, often higher than in other months in the year. *Sell in May and go away* warns investors to sell their stock holdings in May to avoid a seasonal decline on the capital market. For all investment situations, the models (7) and (8) are formulated and solved. In both (special) time periods, the results about the portfolios are similar as in the previous case. The portfolios are totally different (besides some cases with bond funds). The highest return is achieved through variance, semivariance, standard deviation or average absolute negative deviation. The effect of risk measure on the portfolio composition is proved again. For instance, in January the best return is achieved by the average absolute negative deviation risk concept for all funds, as well as for bond funds. On the other side, the portfolio of stock funds composed using this type of risk measure has the second worst level of return. This example illustrates a general fact that any type of risk measure does not bring the highest return in all/most cases.

4 Conclusion

This article primarily searches a dependency of a portfolio composition on the type of used risk measure. To meet this aim, an appropriate methodological approach is proposed. Fuzzy specification of return is practically proved good. The portfolio composition diversity confirms a significant effect of an applied risk measure on the portfolio composition. Another interesting finding is that no risk measure is able to bring the highest/higher return than the others (over all groups of funds and investment time periods). Such an ability of risk measure logically holds only for partial cases. Under these conditions, I think an essential reason for a risk measure selection should be a comprehensibility for a particular user (potential investor). The advantage of the proposed procedure analyzing a risk effect is its applicability for a wider range of investment instruments (stocks, bonds, commodities, etc.). To include them (incl. other open unit trusts) would mean the effect could be confirmed/disproved on a more general level. Further, a risk could be also comprehended as an instable element. Then proposed fuzzy mean-risk model could be modified by fuzzy constraints, or constraints with fuzzy elements.

Acknowledgements

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

References

[1] Alexander, G. J. & Francis, J. C. (1986). *Portfolio analysis*. Englewood Cliffs: Prentice-Hall.
[2] Bellman, R. E. & Zadeh, L. A. (1970). Decision making in a fuzzy environment. *Management Science, 17*, 141–164.
[3] Borovička, A. (2016). Investment portfolio making under uncertainty by stochastic and fuzzy approach. In: *30th European Simulation and Modelling Conference – ESM'2016* (pp. 151–159). Grand Canaria: Eurosis.
[4] Borovička, A. (2014). McCahone's approach and its modifications for a comparison of the fuzzy numbers. In: *Quantitative Methods in Economics (multiple Criteria Decision Making XVII)* (pp.7–13). Bratislava: EKONÓM.
[5] Committee of European Securities Regulators. (2009). *Methodology for the calculation of the synthetic risk and reward indicator*. [online] Available at: http://ec.europa.eu/internal_market/investment/docs/legal_texts/framework/091221-methodilogies-1_en.pdf, [Accessed 5 March 2018].
[6] Investment center. (2018). Funds. [online] Available at: https://cz.products.erstegroup.com/Retail/en/index.phtml, [Accessed 1 April 2018].
[7] Konno, H. & Wijayanayake, A. (1999). Mean-absolute deviation portfolio optimization model under transaction costs. *Journal of the Operations Research Society of Japan*, *42*, 422–435.
[8] Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance, 7*, 77–91.
[9] Markowitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments*. New York: John Wiley & Sons, Inc.
[10] Mashayekhi, Z. & Omrani, H. 2016. An integrated multi-objective Markowitz–DEA cross-efficiency model with fuzzy returns for portfolio selection problem. *Applied Soft Computing*, *38*, 1–9.
[11] Rényi, A. (2007). *Probability theory*. New York: Dover Publications.
[12] Zimmermann, H. J. (1978). Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and Systems*, *1*, 45–55.

Cooperative Multi-Depot Vehicle Routing Problem

Zuzana Čičková [1], Dana Figurová [2]

Abstract. This paper is focused on cooperation possibilities in delivery problems, specifically when the partial subproblems could be modelled as Vehicle Routing Problem. It is considered that the distribution system with customers is assigned to the individual depots. It is assumed that the owners of individual depots (players) could cooperate with each other in coalition in order to reach the savings of transport costs as well as the savings from mutual use of vehicles. This cooperation allows to share customer's service and in that way, it could bring the benefits from merging the transport requirements of individual players. The Cooperative Game Theory explores the positive effects of creating a coalition based on binding agreements. The new presented model is based on the Multi-Depot Vehicle Routing Problem. Generally, in the Multi-Depot Vehicle Routing Problem each vehicle starts its route from one of several sets of depots and its route must end at that same depot. In this paper we accept an assumption that vehicles of one player can be used to serve customers of another player in a possible coalition, which ultimately leads to saving the transport costs.
Keywords: Game Theory, Cooperative Game, Vehicle Routing Problem, Multi-Depot Vehicle Routing Problem

JEL Classification: JEL C61, JEL C71
AMS Classification: AMS 91A12, AMS 91A80

1 Introduction

The Game Theory (GT) is a science that basically examines a wide range of decision situations for multiple participants. One way to explore these decisions is to accept the possibility of cooperation between the participants (players). The game is cooperative, if players can create the coalitions respecting the predetermined obligations and thus they can profit from a common approach. In non-cooperative games the players act independently and we do not consider such commitments.

This paper is focused on the possibilities of cooperation in the Vehicle Routing Problem (VRP), which is one of the most famous routing problems with wide practical applicability. Lots of new constraints are added on the route construction and made practical applications of the VRP (see in [9]). Here are some examples: Messenger Problem in [4], in which the same customer can be served by different vehicles if it reduces the overall costs, the Problem of Synchronized Distribution by Pekár et al. in [7], time constrained VRP by Čičková et al. in [3], Problem with Multiple Depots solved by Karakati and Podgorelec [5]. For example, Ray et al. [8] stated a Multi-Depot Logistics Delivery Problem including the depot selection and shared commodity delivery. In this paper, we will discuss the possibility of cooperation in the case of Vehicle Routing Problem, where we optimize the shipment realized by vehicles that are located in the multiple depots, which belong to the individual players. The model is based on idea of Multi Depot Vehicle Routing Problem (MDVRP), where are several depots instead of just one depot and every customer is visited by a vehicle based at one of the several existing depots, while every vehicle route must start and end at the same depot.

Cooperative Games can be characterized as "Theory, which primarily deals with a coalition of players who coordinate their activities to achieve the further benefits" (see in [2]). In general, shipping costs represent a large part of the company's costs. One solution to reduce these costs is cooperation among the logistic companies. Zibaei et al. show [11], that the Cooperative Game Theory (CGT) can be adopted for modelling the cooperation among the companies. Cooperative Games are concerned with distribution of the cooperation benefits when the players cooperate. Most applications of the CGT are in scheduling, cost saving, negotiation and bargaining (see in [1]). A Cooperative Game actually considers that the players may choose to cooperate by forming some coalitions. In the coalitions, the players might be lucky to receive greater benefits than they could gain individually on their own.

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1,Bratislava, 852 35,Slovak Republic, zuzana.cickova@euba.sk
[2] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1,Bratislava, 852 35,Slovak Republic, danafigurova@gmail.com .

2 Cooperative Multi-Depot Vehicle Routing Problem

The model formulation of a cooperative VRP is based on idea of MDVRP. In the MDVRP each vehicle starts its route from one of several sets of depots and its route must end at that same depot. The requirements of each customer can be satisfied by more than one means of transport. Mostly frequent case in this problem is when customer's demand exceeds a vehicle capacity (collective transport). In theory, the homogeneity of the transport park is often assumed, but in practice it is common to use different vehicles with different capacities, different age and other parameters, not only to minimize overall distance, but also to minimize total costs.

We accept the assumption that the owners of individual depots are different subjects (players). Players are able to cooperate with each other and create the coalitions and reduce their shipping costs. We are considering that each player owns only one depot and also has his own customers, but the players' vehicles in coalition can also be used to serve customers assigned to another player in a possible coalition. We will assume that vehicles have to return back to its starting point (depot) after they came out from their depots. It is not guaranteed, that every vehicle in coalition will be used to serve the customers in order to minimize the total shipping costs.

In this part, we formulate the model of MDVRP with assuming the simultaneous use of vehicles in coalition. In this model, we will assume that there are *k* centers (depots) from which the vehicle can start its route (but the vehicle may not be used which means that it even does not have to start its tour) and the number of vehicles in the depots is not limited. However, each center has a different type of vehicle (in each center, there is one type of vehicle that can differ in size from those, which are located in other centers). We require to satisfy *m* customers from these depots.

We will formulate the mathematical model in the full, edged and appreciated graph $\bar{G} = (N, \bar{H})$. Let: $N^{(1)} = \{p_1, p_2, \dots p_k\}$ is the set of nodes representing the central depots (players) and $N^{(2)} = \{c_1, c_2, \dots c_m\}$ is a set of nodes representing the customers and $N = N^{(1)} \cup N^{(2)}$ is the set of all the nodes of the graph. A capacity of a vehicle in individual depots can be labeled as $g_h, h \in N^{(1)}$. Let $\bar{H} \subset NxN$ represents the set of edges e_{ij}, $i, j \in N$ between all nodes *i* and *j*. Each edge e_{ij} being assigned to a real number denoted d_{ij}, also known as a price of the edge e_{ij}. This assignment is the distance between nodes *i* and *j*. Each of the customers located at $i, i \in N^{(2)}$, requires the import of a certain quantity of goods, generally denoted as $q_i, i \in N^{(2)}$. The goal is to determine the vehicle's routes, which will satisfy the requirements of all customers. Customer requirements will only be realized in the whole (if the vehicle serves the customer, its entire delivery requirement will be realized), with no vehicle capacity exceeded. The main objective is to minimize the total traveled distance. In the model we assume implicitly that $q_i \leq \max\{g_h, h \in N^{(1)}\}$ for all $i \in N^{(2)}$ which means, that the size of each customer's requirement will not exceed the capacity of the vehicle.

In general, one way to mathematically describe the routing problems is using binary variables x_{ijh}, $i, j \in N$, $h \in N^{(1)}, i \neq j$ that enable to model if the node *i* precedes node *j* in a route of the vehicle from the *h*-th depot. Further on, the variables u_{ih}, $i \in N^{(2)}, h \in N^{(1)}$ that based on the known Miller-Tucker-Zemlin formulation of Traveling Salesman Problem in [6] are used. Those variables represent cumulative demand of customers on one particular route.

Let us recapitulate the model parameters and model variables more clearly.

Parameters

- $N^{(1)} = \{p_1, p_2, \dots p_k\}$ – set of depots (players),
- $N^{(2)} = \{c_1, c_2, \dots c_m\}$ – set of customers (served nodes),
- $N = N^{(1)} \cup N^{(2)}$ – set of all nodes (customers and depots),
- d_{ij}, $i, j \in N, i \neq j$ – distance moving from node *i* to node *j*,
- q_i, $i \in N^{(2)}$ – demand of *i*-th customer,
- g_h, $h \in N^{(1)}$ – capacity of vehicles in the *h*-th depot,

Variables

- x_{ijh} $i, j \in N, h \in N^{(1)}, i \neq j$ representing if the node *i* precedes node *j* in a final route of vehicle from the *h*-th depot ($x_{ijh} = 1$) or not ($x_{ijh} = 0$).
- $u_{ih}, i \in N^{(2)}, h \in N^{(1)}$ based on Miller - Tucker - Zemlin's formulation but representing vehicle load on its route to the *i*-th node (including).

Model of Multi Depot Vehicle Routing Problem with simultaneous use of vehicles can be written as:

$$c = \min f(\mathbf{X},\mathbf{u}) = \sum_{i\in N}\sum_{\substack{j\in N\\ i\neq j}}\sum_{h\in N^{(1)}} d_{ij}x_{ijh} \tag{1}$$

$$\sum_{i\in N}\sum_{h\in N^{(1)}} x_{ijh} = 1,\ j\in N^{(2)},\ i\neq j \tag{2}$$

$$\sum_{j\in N} x_{ijh} = \sum_{j\in N} x_{jih},\ i\in N^{(2)}, h\in N^{(1)} \tag{3}$$

$$\sum_{j\in N^{(2)}} x_{iji} = \sum_{j\in N^{(2)}} x_{jii},\ i\in N^{(1)} \tag{4}$$

$$\sum_{j\in N^{(2)}} x_{iji} = \sum_{j\in N^{(2)}}\sum_{h\in N^{(1)}} x_{ijh},\ i\in N^{(1)} \tag{5}$$

$$u_{ih} + q_j - g_h(1-x_{ijh}) \le u_{jh},\ i\in N,\ j\in N^{(2)},\ h\in N^{(1)},\ i\neq j \tag{6}$$

$$u_{ii} + q_j - g_i(1-x_{iji}) \le u_{ji},\ i\in N^{(1)},\ j\in N^{(2)},\ i\neq j \tag{7}$$

Constraint set (2) guarantees that each customer will be visited exactly once and exactly by one vehicle. Conditions (3) a (4) ensure the balance of the route. Constraint set (5) provides the balance of the number of routes from the depot (if this depot is used). Constraints (6) a (7) are the sub-tour elimination conditions. The scalar *c* (1) represents the minimum value of the total travelled distance.

Now we accept the assumption that the owners of individual depots are different subjects (players). Players are able to cooperate with each other and create the coalitions and reduce their shipping costs. We are considering that each player owns only one depot and also has his own customers, but the player's vehicles in coalition can also be used to serve customers assigned to another player in a possible coalition. The number of possible coalitions is $2^k - 1$ (where *k* represents the number of players). Let $N_S^{(1)} \subset N^{(1)}$ be a set of players' coalitions. We will also divide a set of customers based on their membership to individual players, then the customers of coalition *S* will be labeled as $N_S^{(2)}$. Thus, the set $N_S = N_S^{(1)} \cup N_S^{(2)}$ is the set representing the depots and the customers of the coalition *S*. Then, the optimal distributions of individual coalitions can be quantified by using the aforementioned model, when $N = N_S, N^{(1)} = N_S^{(1)}, N^{(2)} = N_S^{(2)}$.

3 Results

Firstly, we are solving our modified model of cooperative MDVRP (1) – (7). The data were obtained from website (see at [10]), where we chose the first instance of asymmetric Travelling Salesman Problem (br17.atsp). We are working with the distance matrix, which we obtained from origin data by Floyd algorithm using Microsoft Excel.

Consider the net of 17 nodes. We will assume that there are *4* depots from which the vehicle can start its route (but the vehicle may not be used, it means that it even does not have to start its tour). So, we consider the distribution problem with multiple depots, whereby we have four suppliers to serve the certain customers. Suppliers or players (owners of individual depots) are expressed as $p_k = \{p_1, p_2, p_3, p_4; k = 1,2,3,4\}$. Each player owns one depot with one vehicle. The different capacity of each vehicle is given by $g_{p_1} = 250$, $g_{p_2} = 220$, $g_{p_3} = 210$ *and* $g_{p_4} = 200$. Customers, who are strictly assigned to the individual depots (players), will be marked as: $\{c_1, c_2, c_3\}\ for\ p_1$, $\{c_4, c_5, c_6\}\ for\ p_2$, $\{c_7, c_8, c_9\}\ for\ p_3, \{c_{10}, c_{11}, c_{12}, c_{13}\}\ for\ p_4$. In the case of the creation the coalition $S \subseteq N^{(1)}$ we know that there are exactly 15 possible coalitions between the players.

We solve the cooperative MDVRP by using the model (1) - (7) for the created coalitions *S*: {1,2}}, {1,4}, {2,3}, {2,4}, {3,4}, {1,2,3}, {1,2,4}, {1,3,4} {2,3,4 } and {1,2,3,4} by GAMS software. To obtain the optimal solution, we used the solver Cplex 12.2.0.0 on the personal computer INTEL® Core ™ 2 CPU, E8500 @ 3.16 GB RAM for Windows 10.

Our interest is to find optimal solutions by using the modified model of cooperative MDVRP (1) - (7). Our main idea is to prove that there is a reduction in total shipping costs through mutual cooperation between suppliers.

Table 1 presents the total transport costs *TC*(*S*) of individual coalitions obtained from our model MDVRP with assumptions mentioned above. It means that every vehicle has to return back into their starting point (depot) and

it is possible that not every player's vehicle will be used in coalition. After this, we also compare it with results, which we obtained with one extra assumption:

$$\sum_{j \in N^{(2)}} x_{iji} \geq 1 \tag{8}$$

This assumption (8) indicates that from every single depot in coalition must come out at least one vehicle. We assume unlimited number of vehicles. We can also see the optimal tours for individual coalitions.

Coalition	Optimal route	Total cost		Total cost with new assumption
{1,2}	p_1-c_3-c_2-c_1-c_5-c_4-p_1 p_2-c_6-p_2	31	=	31
{1,3}	p_1-c_1-c_3-c_2-c_9-c_7-p_1, p_1-c_8-p_1	31	<	34
{1,4}	p_1-c_{13}-c_{11}-c_3-c_2-c_{12}-c_{10}-p_1 p_4-c_1-p_4	29	=	29
{2,3}	p_2-c_5-c_4-c_8-p_2, p_2-c_6-p_2, p_2-c_7-p_2, p_2-c_9-p_2	13	<	18
{2,4}	p_2-c_6-p_2, p_2-c_{10}-p_2, p_2-c_{11}-c_{12}-c_5-c_{13}-c_4-p_2	27	<	28
{3,4}	p_3-c_8-c_{11}-c_{12}-c_{13}-c_7-c_9-p_3, p_3-c_{10}-p_3	29	<	30
{1,2,3}	p_1-c_3-c_2-c_1-c_5-c_4-p_1, p_1-c_8-p_1 p_2-c_6-p_2, p_2-c_7-p_2, p_2-c_9-p_2	31	<	37
{1,2,4}	p_1-c_{11}-c_3-c_{12}-c_2-c_5-c_{13}-c_4-p_1 p_2-c_6-p_2, p_4-c_1-p_4	27	=	27
{1,3,4}	p_1-c_8-p_1, p_1-c_{11}-c_3-c_{12}-c_2-c_{13}-c_7-c_9-p_1 p_3-c_{10}-p_3, p_4-c_1-p_4	24	=	24
{2,3,4}	p_2-c_6-p_2, p_2-c_7-p_2, p_2-c_9-p_2, p_2-c_{11}-c_{12}-c_5-c_{13}-c_4-c_8-p_2 p_3-c_{10}-p_3	24	<	25
{1,2,3,4}	p_1-c_8-p_1 p_2-c_3-c_{11}-c_{12}-c_2-c_5-c_{13}-c_4-p_2, p_2-c_6-p_2 p_2-c_7-p_2, p_2-c_9-p_2 p_3-c_{10}-p_3 p_4-c_1-p_4	21	=	21

Table 1 Optimal routes in our Multi-Depot Vehicle Routing Problem

We can summarize various results from Table 1. If player 1 cooperates with players 2 and 3, their minimum total cost is 31 units. For the comparison, if we obtained the new assumption (from every single depot in coalition must come out at least one vehicle), coalition $\{1,2,3\}$ has to pay about 6 units of shipping cost more. So, we can state that if we accept the assumption about not using every vehicle in coalitions, we obtained much better results. We can see the difference in our results by comparing the fourth and fifth column in Table 1.

4 Conclusion

In this paper, we focused on the cooperative MDVRP with assumption of cooperation between the players to minimize the total shipment costs. Our main task was to compare the results obtained by solving our model (1)-(7) and the results solved by this model with adding the extra assumption that every vehicle in coalition has to come out from its depot and has to be used in optimal route. Our main idea was to prove that there is a reduction in total transport costs. In one hand, there is a cost reduction through mutual cooperation between individual suppliers and on the other hand trough acceptance of different assumptions. By comparing our results, we have taken the decision that our model without the extra assumption has produced much better results. In conclusion, we can say that after the comparison our models with and without the extra assumption, the results of the optimal paths of each player's vehicles have improved or stay at the same level of shipment costs.

Acknowledgements

This work was supported by the project VEGA 1/0351/17 Application of selected models of Game Theory in economic problems of Slovakia.

References

[1] Barron, E. N. (2013). *Game Theory: An introduction.* New York: John Wiley & Sons Inc.
[2] Branzei, R., Dimtrov, D., & Tijs, S. (2008). *Models in cooperative game theory*. New York: Springer Science and Business Media.
[3] Čičková, Z., Brezina, I., & Pekár J. (2013). Solving the real-life vehicle routing problem with time windows using self organizing migrating algorithm. *Journal for economic theory, economic policy, social and economic forecasting,* 61, 497-513.
[4] Fábry, J.(2015). Insertion method for multiple messenger problem with multiple depots. *Communications*, 17, 11-14.
[5] Karakatič, S. & Podgorelec, V. (2015). A survey of genetic algorithms for solving multi depot vehicle routing problem. *Applied Soft Computing*, 27, 519-532.
[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] Pekár, J., Brezina, I., & Čičková, Z. (2017). Synchronization of capacitated vehicle routing problem among periods. *Journal for economic theory, economic policy, social and economic forecasting,* 65, 66-78.
[8] Ray, S., Soeanu, A., Berger, J. & Debbabi, M. (2014). The multi-depot split-delivery vehicle routing problem: Model and solution algorithm. *Knowledge-Based Systems*, 71, 238-265.
[9] Toth,P. & Vigo, D. 2002. *The Vehicle Routing Problem,* Montreal: Society for Industrial and Applied Mathematics Press.
[10] TSPLIB, (2018). Database – TSPLIB. Available at: http://elib.zib.de/pub/mp-testdata/tsp/tsplib/atsp/br17.atsp [Accessed 15 Apr. 2018].
[11] Zibaei, S., Hafezalkotob. A., and Ghashami, S. Cooperative vehicle routing problem: an opportunity for cost saving. *Journal of Industrial Engineering International,* 12, 271-286.

Cyclicality of Job-Finding and Separation Rates in Europe

Ondřej Čížek[1]

Abstract. Standard Shimer's [9] methodology of measuring job-finding and separation rates is applied and these measures are calculated for many individual European labor markets using commonly available data. Relation of these transition rates to business cycle is investigated by means of regression and correlation analysis. An important finding is that separation rate is counter-cyclical while a job-finding rate is rather acyclical for most of the European countries which is interesting as it is in line with commonly believed facts that were criticized by Shimer for the US economy.

Keywords: unemployment, job-finding rate, separation rate.

JEL Classification: E24, J64
AMS Classification: 91G70

1 Introduction

European labor markets are influenced by many factors (Dimian et al. [2]). This paper focuses on job-finding and separation probability. Specifically, Shimer's [10] measure of these transition probabilities is applied in this paper in order to study European labor markets. Empirical analysis is performed not to Europe as a whole but for many individual European countries due to its heterogeneity (Hušek, Formánek [5]). Similar labor market analysis was performed by Čabla, Malá [1], Flek, Mysíková [3] or Maleček [6]. Empirical analysis of European labor markets not on a country-level but on a level of individual regions was performed by Formánek, Hušek [4] and Střeleček et al. [12]. The main goal of the paper is to analyze a relation of transition rates to the business cycle for many individual European economies and compare the results with Shimer's conclusions for the US economy.

The structure of the paper is as follows. Firstly, Shimer's methodology of measuring transition rates is described in chapter 2. Data are presented in section 3. Empirical analysis of cyclicality of job-finding and separation rates is performed in chapter 4 and final chapter 5 concludes.

2 Job-finding and separation rates

Shimer's [10] methodology is adopted in modeling unemployment dynamics. The model of unemployment is in continuous time, but data are measured at discrete dates. The time interval $[t,t+1)$ is referred to as the period t. It is assumed that during period t, all unemployed workers find a job according to a Poisson process with arrival rate $f_t \equiv -\ln(1-F_t) \geq 0$, where F_t is the corresponding job finding probability.[2] Similarly, all employed workers lose their job according to a Poisson process with arrival rate $s_t \equiv -\ln(1-S_t) \geq 0$, where S_t is the corresponding employment exit probability.

The first goal is to calculate F_t and S_t from commonly available data. Short-term unemployment will be introduced in order to calculate F_t. Specifically, $U_t^s(\tau)$ will represent the number of workers who are unemployed at time $t+\tau$ but were employed at some time $t' \in [t,t+\tau]$, $\tau \in [0,1]$. Unemployment and short-term unemployment evolves according to

$$\dot{U}(t+\tau) = E(t+\tau)\cdot s_t - U(t+\tau)\cdot f_t, \tag{1}$$

$$\dot{U}_t^s(\tau) = E(t+\tau)\cdot s_t - U_t^s(\tau)\cdot f_t, \tag{2}$$

where $E(t+\tau)$, $U(t+\tau)$ represents number of employed and unemployed workers at time instant $t+\tau$. Note that t is fixed in these equations and that $\tau \in [0,1]$.

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 130 67, Prague 3, Czech Republic, cizeko@vse.cz.

[2] The difference between the rate and the corresponding probability is that the probability relates to the discrete time while the rate to the continuous time.

To solve for F_t, eliminate $E(t+\tau)\cdot s_t$ between these two equations, which gives $\dot{U}(t+\tau)-\dot{U}_t^s(\tau)=-\left(U(t+\tau)-U_t^s(\tau)\right)\cdot f_t$. Solution to this differential equation is given by $U(t+\tau)-U_t^s(\tau)=K\cdot e^{-f_t\cdot\tau}$. Because $U_t^s(0)=0$ from definition, we have $U(t)=K$. For $\tau=1$ we obtain $U(t+1)-U_t^s(1)=U(t)\cdot e^{-f_t}$, or[3] $U_{t+1}-U_t^s(1)=U_t\cdot(1-F_t)$. Simple algebraic manipulation leads to

$$F_t=1-\frac{U_{t+1}-U_t^s(1)}{U_t}. \tag{3}$$

This equation serves to calculate F_t using statistical data for unemployment U_t and short-term unemployed $U_t^s(1)$. The job-finding rate is then calculated as $f_t=-\ln(1-F_t)$.

Furthermore, solving differential equation (1) leads to

$$U_{t+1}=\frac{s_t}{s_t+f_t}\cdot\left(1-e^{-(s_t+f_t)}\right)\cdot L_t+e^{-(s_t+f_t)}\cdot U_t, \tag{4}$$

where $L_t\equiv E_t+U_t$ represent the labor force. Job separation rate s_t is obtained from solving the equation (4) numerically and the corresponding probability is given by $S_t=1-e^{-s_t}$.

3 Data

The data for individual European countries were obtained from the Eurostat database.[4] Seasonal adjustment was performed in Eviews 8 (method Census X–12) in cases where the adjustment was not performed by the Eurostat. The main variables are job-finding and separation rates. These data are available for Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Netherlands, Poland, Portugal, Slovenia, Spain, Sweden and United Kingdom. Job-finding and separation rates were calculated by the Shimer's [10] method described earlier in chapter 2. Empirical counterpart to the variable $U_t^s(1)$ is the number of workers unemployed at the beginning of the quarter t, whose unemployment has not exceeded one quarter (three months).[5] GDP in chain-linked volumes for the same European countries was obtained as well in order to empirically investigate a relation of job-finding and separation rates to a business cycle.

4 Relation of transition rates to business cycle

It is common in empirical literature to measure business cycle as a cyclical component of GDP which is most often estimated by Hodrick-Prescott filter. It turned out that job-finding and separation rates were only very slightly correlated with this traditional measure of the output gap. For this reason, GDP growth was taken as a measure of a business cycle in this paper. Specifically, GDP growth is calculated according to $\Delta y_t=(y_t-y_{t-1})/y_{t-1}$, where y_t is GDP in time t.

A relation of transition rates to the business cycle Δy_t is investigated by two methods. The first method is based on regression equations and the second methods is based on a correlation analysis. Job-finding and separation rates will be modeled by first-order autoregressive processes because of their inertial properties. Specifically, the following regression equations are assumed:

$$s_t=\alpha_0+\alpha_1\cdot s_{t-1}+\alpha_2\cdot\Delta y_t+\varepsilon_t^s, \tag{5}$$

$$f_t=\beta_0+\beta_1\cdot f_{t-1}+\beta_2\cdot\Delta y_t+\varepsilon_t^f, \tag{6}$$

where α_0, $\beta_0>0$, α_1, $\beta_1\in(0,1)$, $\alpha_2<0$, $\beta_2>0$ and ε_t^s, ε_t^f are i.i.d. random errors. Regression equations (5) and (6) were estimated by ordinary least squares and relation between transition rates and business cycle were analyzed by testing standard hypothesis $H_0:\alpha_2=0$ and $H_0:\beta_2=0$. The second method of analyzing relation between transition rates and business cycle is by means of correlation coefficients $corr(s_t,\Delta y_t)$, $corr(f_t,\Delta y_t)$

[3] Using the convention to denote time in the lower subscript when measured in discrete time.
[4] http://ec.europa.eu/eurostat/data/database
[5] The name of this time series in the Eurostat database is 'Unemployment by sex, age and duration of unemployment [lfsq_ugad]'.

and by testing $H_0 : corr(s_t, \Delta y_t) = 0$, $H_0 : corr(f_t, \Delta y_t) = 0$. The results for separation rate are summarized in the following table.[6]

		separation rate				
		autoregressive equation				correlation
country	sample	α_0	α_1	α_2	R^2	$corr(s_t, \Delta y_t)$
Cyprus	2004Q1-2016Q3	0.007 (0.00)	0.77 (0.00)	-0.16 (0.02)	0.80	-0.64 (0.00)
Czech Republic	1998Q1-2016Q3	0.004 (0.00)	0.77 (0.00)	-0.12 (0.00)	0.84	-0.57 (0.00)
Finland	1998Q1-2016Q3	0.03 (0.00)	0.49 (0.00)	-0.13 (0.00)	0.43	-0.44 (0.00)
Germany	2005Q1-2016Q3	0.002 (0.03)	0.90 (0.00)	-0.04 (0.01)	0.92	-0.18 (0.22)
Greece	1998Q2-2016Q4	0.003 (0.01)	0.88 (0.00)	-0.06 (0.02)	0.89	-0.59 (0.00)
Hungary	1998Q4-2016Q3	0.003 (0.01)	0.85 (0.00)	-0.06 (0.02)	0.81	-0.38 (0.00)
Ireland	2005Q4-2016Q3	0.004 (0.03)	0.86 (0.00)	-0.02 (0.11)	0.84	-0.31 (0.04)
Latvia	2006Q4-2016Q4	0.01 (0.00)	0.55 (0.00)	-0.18 (0.00)	0.64	-0.63 (0.00)
Lithuania	2007Q4-2016Q4	0.01 (0.00)	0.67 (0.00)	-0.11 (0.00)	0.71	-0.57 (0.00)
Netherlands	2002Q1-2016Q3	0.001 (0.04)	0.94 (0.00)	-0.04 (0.11)	0.92	-0.13 (0.32)
Poland	2002Q2-2016Q4	0.004 (0.03)	0.84 (0.00)	-0.04 (0.22)	0.69	-0.09 (0.49)
Portugal	2000Q2-2016Q4	0.002 (0.052)	0.91 (0.00)	-0.02 (0.71)	0.89	-0.38 (0.00)
Slovenia	2001Q3-2016Q4	0.005 (0.00)	0.68 (0.00)	-0.04 (0.14)	0.58	-0.49 (0.00)
Spain	1997Q4-2016Q4	0.006 (0.053)	0.92 (0.00)	-0.12 (0.13)	0.94	-0.75 (0.00)
Sweden	2006Q1-2016Q4	0.01 (0.01)	0.69 (0.00)	-0.04 (0.29)	0.53	-0.26 (0.09)
United Kingdom	1999Q1-2016Q3	0.007 (0.00)	0.77 (0.00)	-0.08 (0.01)	0.69	-0.37 (0.00)

Table 1: Relation between separation rate and growth of the GDP

The results show that separation rate in countercyclical and this result is quite robust. The coefficient α_2 has expected negative sign for all economies and the correlation $corr(s_t, \Delta y_t)$ turned out to be negative for all economies as well. Nonetheless, the hypothesis $H_0 : \alpha_2 = 0$ is not rejected in all cases using the conventional 5 % level of significance. It is not rejected for Ireland, Netherlands, Poland, Portugal, Slovenia, Spain and Sweden. The hypothesis $H_0 : corr(s_t, \Delta y_t) = 0$ is not rejected for some countries as well, but this is the case only for Ger-

[6] P-values relating to t-tests of the estimated coefficients are indicated in parentheses below the parameters.

many, Netherlands and Poland. Similar analysis was performed also for a job-finding rate and the results are summarized in the following table.

		job-finding rate				
		autoregressive equation				correlation
country	sample	β_0	β_1	β_2	R^2	$corr(f_t, \Delta y_t)$
Cyprus	2004Q1-2016Q3	0.17 (0.00)	0.46 (0.00)	3.89 (0.03)	0.38	0.40 (0.00)
Czech Republic	1998Q1-2016Q3	0.06 (0.00)	0.73 (0.00)	-0.22 (0.66)	0.47	-0.02 (0.89)
Finland	1998Q1-2016Q3	0.16 (0.00)	0.72 (0.00)	-0.67 (0.11)	0.64	-0.37 (0.00)
Germany	2005Q1-2016Q3	0.02 (0.19)	0.93 (0.00)	0.40 (0.14)	0.85	0.10 (0.51)
Greece	1998Q2-2016Q4	0.03 (0.01)	0.77 (0.00)	0.04 (0.88)	0.60	0.23 (0.049)
Hungary	1998Q4-2016Q3	0.03 (0.047)	0.85 (0.00)	0.26 (0.55)	0.68	0.23 (0.053)
Ireland	2005Q4-2016Q3	0.06 (0.02)	0.71 (0.00)	0.24 (0.28)	0.51	0.03 (0.84)
Latvia	2006Q4-2016Q4	0.15 (0.00)	0.38 (0.01)	1.73 (0.048)	0.31	0.45 (0.00)
Lithuania	2007Q4-2016Q4	0.17 (0.00)	0.29 (0.03)	0.16 (0.81)	0.15	0.12 (0.48)
Netherlands	2002Q1-2016Q3	0.06 (0.02)	0.78 (0.00)	1.09 (0.08)	0.60	0.13 (0.34)
Poland	2002Q2-2016Q4	0.05 (0.03)	0.84 (0.00)	-0.29 (0.67)	0.67	0.002 (0.99)
Portugal	2000Q2-2016Q4	0.13 (0.00)	0.34 (0.00)	1.50 (0.03)	0.21	0.32 (0.01)
Slovenia	2001Q3-2016Q4	0.13 (0.00)	0.32 (0.01)	1.27 (0.08)	0.14	0.19 (0.13)
Spain	1997Q4-2016Q4	0.03 (0.09)	0.90 (0.00)	1.82 (0.02)	0.89	0.42 (0.00)
Sweden	2006Q1-2016Q4	0.09 (0.09)	0.83 (0.00)	1.48 (0.02)	0.70	0.10 (0.51)
United Kingdom	1999Q1-2016Q3	0.03 (0.10)	0.92 (0.00)	0.90 (0.19)	0.88	0.33 (0.01)

Table 2: Relation between job-finding rate and growth of the GDP

The table shows that there is some evidence for a pro-cycle behavior of job-finding rate, but this evidence is rather weak. The coefficient β_2 is not positive in all cases – it is negative for the Czech Republic, Finland and Poland. Similarly, the correlation $corr(f_t, \Delta y_t)$ turned out to negative for the Czech Republic and Finland. Moreover, statistical hypothesis $H_0 : \beta_2 = 0$ is rejected only for Cyprus, Latvia, Portugal, Spain and Sweden. Similarly, the null hypothesis $H_0 : corr(f_t, \Delta y_t) = 0$ of acyclical behavior of job-finding rate is rejected only for half of the studied countries – Cyprus, Finland, Greece, Latvia, Portugal, Spain and United Kingdom.

These results are not fully in line with the conclusions of previous studies. Specifically, Shimer [9] argues that commonly believed fact that a job-finding rate is acyclical and a separation rate is counter-cyclical does not hold and brings evidence that a job-finding rate was pro-cyclical while a separation rate was nearly acyclical in the United States from 1948 to 2004. My results evidently do not support Shimer's arguments and are in line with the above mentioned commonly believed fact that a job-finding rate is acyclical and a separation rate is counter-cyclical for most of the European countries. This result was obtained by applying the same methodology as was applied by Shimer and thus the difference is not caused by different methodologies.

5 Conclusion

The aim of the paper was to investigate a relation of transition rates measured by the Shimer's [10] methodology to business cycle. Empirical analysis was performed on a very large set of European countries. Separation rate turned out to be counter-cyclical while a job-finding rate seems to be rather acyclical. These results support the commonly believed facts about the cyclicality of transition rates, but are in contrast with Shimer's findings for the US economy which is interesting as Shimer's methodology was applied in this paper.

The analysis could be extended in a variety of ways. Alternative functional forms of regression models could be taken into account, dynamical properties might be studied or possible instability due to the current crisis might also be modeled. Usefulness of information contained in the transition rates for labor market analysis could also be studied as discussed by Pánková [7], [8]. Detailed analysis along these lines might be a topic for future research.

Acknowledgements

This work was processed with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague (IP400040) and it was supported by the Internal Grant Agency of University of Economics, Prague under Grant F4/21/2018.

References

[1] Čabla, A., Malá, I. (2017). Využití metod analýzy přežití pro modelování doby nezaměstnanosti v České republice. *Politická ekonomie*, **65**(4), 501–519.
[2] Dimian, G.C., Ileanu, B., Jablonský, J., Fábry, J. (2013). Analysis of European Labour Market in the Crisis Context. *Prague Economic Papers*, **22**(1), 50–71.
[3] Flek, V., Mysíková, M. (2015). Unemployment Dynamics in Central Europe: A Labour Flow Approach. *Prague Economic Papers*, **24**(1), 73–87.
[4] Formánek, T., Hušek, R. (2017). Spatial Panel Data Models – Stability Analysis with Application to Regional Unemployment. In Pražák, P. (Ed.), *Mathematical Methods in Economics (MME 2017)* (pp. 183–188). Hradec Králové: University of Hradec Králové, ISBN 978-80-7435-678-0.
[5] Hušek, R, Formánek, T. (2011). Srovnání konvergence ekonomiky ČR a vybraných zemí eurozény na základě analýzy funkcí odezvy a nabídkových či poptávkových šoků. *Politická ekonomie*, **59**(3), 291–309.
[6] Maleček, P. (2014). Toky dlouhodobé nezaměstnanosti. *Politická ekonomie*, **62**(4), 560–576.
[7] Pánková, V. (1997). Modelování dynamických procesů – problem dat. *Politická ekonomie*, **45**(4), 592–602.
[8] Pánková, V. (2002). Předpověď jako ukazatel kvality ekonometrického modelu. *Acta Oeconomica Pragensia*, **10**(7), 63–69.
[9] Shimer, R. (2005). The Cyclicality of Hires, Separations, and Job-to-Job Transitions. Federal Reserve Bank of St. Louis *Review*, July/August 2005, 493-508.
[10] Shimer, R. (2012). Reassessing the Ins and Outs of Unemployment. *Review of Economic Dynamics*, **15**(2), 127–148.
[11] Simionescu, M. (2015). The Improvement of Unemployment Rate Predictions Accuracy. *Prague Economic Papers*, **24**(3), 274–286.
[12] Střeleček, F, Zdeněk, R., Lososová, J. (2010). Vývoj zaměstnanosti v znevýhodněných oblastech v letech 2002–2006. *Politická ekonomie*, **58**(6), 761–773.

Measuring the Changes in Resource Capacity and Resource Distribution: A Health Care Example

Martin Dlouhý[1]

Abstract. The objective of this study is to measure and evaluate development of the capacity levels of multiple resources and the geographic inequality in the resource distribution by one method. The proposed method is based on the data envelopment analysis that is able to transform multiple resources into a single virtual resource. The capacity levels are compared by the ratio of total virtual resources and the geographic inequality is measured by the Gini coefficient and the Robin Hood Index. The proposed method is illustratively applied to the physicians and nurses in the Czech Republic. The changes in the resource capacities and in the resource distribution are compared for years 2005 and 2015. We observe an improvement in both the resource capacity and the geographic distribution of both physicians and nurses between the years 2005 and 2015.

Keywords: Geographic Inequality, Gini Coefficient, Robin Hood Index, Data Envelopment Analysis.

JEL Classification: D63 C61
AMS Classification: 90C90

1 Introduction

To ensure sufficient capacities and equal distribution of public resources is a usual objective of public policy. For example in health care, access to health services and equal access to them for all, irrespectively of social status, ethnicity or place of living, are important objectives of the national health policy. The importance of this objective affects the overall organization of the national health system. On the other hand, the market mechanism allocates resources according to the willingness and ability to pay. As a consequence a supply of health services will be concentrated in rich areas, whereas poor areas, albeit being usually those with the greatest health needs, will not be served adequately. However, European health systems are mostly publicly funded, so the insufficient resource capacities or unequal regional distribution of health resources can be considered as a failure.

Let us use the Czech Republic as an illustrative example. In 2015, the number of physicians per 10,000 inhabitants was 36.30, which is an increase in comparison to 33.46 physicians in the year 2005. In 2015, the number of nurses per 10,000 inhabitants was 77.48, which is a decrease in comparison to 80.37 nurses in the year 2005. It is assumed that substitution between physicians and nurses is possible. In a region with more physicians, a lower number of nurses is needed, and vice versa, in a region with more nurses, a lower number of physicians is needed. In this case, in which the number of physicians increased and the number of nurses decreased, it is not clear whether the overall access to health resources improved or not. One of the objectives of this contribution is to develop a measure that is able to resolve this question.

We also want to know how these health resources, physicians and nurses, are distributed geographically. So we need to have an inequality measure that is able to work with multiple resources. The objective of this study is to measure and evaluate development of the level of resource capacity and the geographic inequality in the resource distribution by one method, which extends our previous research [7, 8]. The proposed method will be illustratively applied to the physicians and nurses in the Czech Republic.

2 Method

2.1 Inequality Measurement

Measures of inequality express the variation in observed variable by a single number. There is a variety of inequality measures described in the literature [6, 10, 11] and it is not possible to say that one inequality measure is better than the other. The simple measures of inequality are the ranges, decile ratios, coefficient of variation, etc.

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, 4 Winston Churchill Sq., Prague, Czech Republic, e-mail: dlouhy@vse.cz.

The popular measure of inequality is the Gini coefficient that is derived from the Lorenz curve, a cumulative frequency curve that compares the empirical distribution of the observed variable with the uniform distribution that represents perfect equality. The Gini coefficient ranges between 0, which occurs in the case of perfect equality, and 1, which occurs in the case of perfect inequality. The Gini coefficient (GC) can be calculated by the formula (1):

$$GC = \frac{1}{2r}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i\pi_j|r_i - r_j|\right) \tag{1}$$

where π_i is the population proportion, r_i is the number of resource per capita, r is the average value of the resource per capita, and n is the number of geographic areas.

The Robin Hood Index (RHI) 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 Robin Hood Index is calculated by the formula (2):

$$RHI = \frac{1}{2}\sum_{i=1}^{n}|\pi_i - \rho_i| \tag{2}$$

where π_i is the population proportion, ρ_i is the resource proportion, and n is the number of geographic areas. The Robin Hood Index is multiplied by 100 to be in percentages.

Let us assume that we measure inequality in geographic distribution in the case of multiple health resources. The health resources as physicians and nurses are, at least to some extent, substitutes. Hence the region with fewer physicians may compensate such disadvantage by a larger number of nurses. In such a case, the multiple resource inequality is lower than expected from separate inequality measurements. To cope with cases with multiple health resources, one can use multiple criteria decision making for setting relative resource weights. The question is how to obtain such weights. More flexible approaches to estimate resource weights and the marginal rate of technical substitution can be based on the production function. Health resources are inputs and population (as a measure of health need) is the single output. The production function can be estimated by econometric methods (e.g., stochastic frontier analysis) or by data envelopment analysis that is described in section 2.2.

2.2 Data Envelopment Analysis

Data envelopment analysis (DEA) constructs the production frontier and evaluates the technical efficiency of production units. DEA is a based on the theory of mathematical programming that estimates the production frontier as the piecewise linear envelopment of the observed data [2]. A variety of DEA models with many extensions and modifications has been developed that can be found in textbooks that also present many examples of applications from both the private and public sectors [3, 4, 5, 9].

The production unit uses a number of inputs to produce a number of outputs. The 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. Each unit can choose its own weights of inputs and outputs in order to maximize its efficiency score. A technically efficient 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 (the output maximization DEA model) or, alternatively, the minimum amounts of inputs required to pro-duce the given amount of output (the input minimization DEA model).

Suppose that we have a sample with n production units that use m inputs to produce r outputs. The mathematical formulation of the input-oriented version of the constant returns-to-scale DEA model for production unit q is:

$$\begin{aligned}
&Maximize && \varphi_q = \sum_{k=1}^{r} u_k y_{kq} \\
&subject\ to && \sum_{k=1}^{r} u_k y_{kj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1, 2, \dots, n, \\
& && \sum_{i=1}^{m} v_i x_{iq} = 1, \\
& && u_k \ge \varepsilon, \quad k = 1, 2, \dots, r, \\
& && v_i \ge \varepsilon, \quad i = 1, 2, \dots, m,
\end{aligned} \tag{3}$$

where φ_q is the technical efficiency score, x_{ij} is the amount of input i used by unit j, y_{kj} is the amount of output k produced by unit j, and ε represents an infinitesimal constant. The output weights u_k and input weights v_i are variables in the model. In the input-oriented model, the efficiency score φ_q is one if the unit q is technically efficient, and is lower than one if the unit is technically inefficient. The efficiency score measures a size of input reduction that makes production unit q technically efficient.

The DEA model (3) has to be solved for each unit. The input data was normalized by dividing original values by the maximum value for each input and output. The value of ε was set to 0.00000001, which is sufficient to ensure feasibility and boundedness [1]. The calculations were made in *DEA-Excel Solver 2014*, which is a MS Excel-based application for solving DEA models. It is available at https://webhosting.vse.cz/jablon/.

2.3 Capacity and Inequality Measurement

Suppose a situation with two resources (inputs) and one output (the regional population that serves as an estimation of health need). Let us have regions A, B, and C. Regions A and B lie on the production frontier are technically efficient, and region C is inefficient, having the input-oriented efficiency score 0.8, which was estimated by data envelopment analysis. A lower level of inefficiency in this situation represents a higher level of total resources that is available for regional population. The efficiency score of the input-oriented constant returns-to-scale model, which is lower than one, expresses the excess of resources above the most badly served regions that are represented by the set of DEA efficient units. This means that it is better for you to live in inefficient regions! The efficiency score 0.8 means that there is a possibility of 20% resource reduction in the given region or that the given combination of health resources is able to serve a 25% larger population (1/0.8 = 1.25). By the efficiency scores for regions, multiple health resources are transformed into a single virtual resource the value of which is calculated as a regional population (or a proportion of regional population) multiplied by reciprocal value of the efficiency score.

The changes in resource capacity levels are compared by the ratio of total virtual resources from different time periods. It is assumed that over the time period covered by a panel there is no technology change. This assumption is described by the so-called intertemporal production frontier [12], which is estimated from the observations throughout the whole observation period.

The following method that is able to measure both the resource capacity and inequality distribution is proposed:

1. Construct intertemporal production frontier. For each region from all time periods, calculate the efficiency scores φ_{it} by the input-oriented version of the constant returns-to-scale DEA model with health resources as inputs and the regional population (or any other measure of health need) as a single output.
2. Calculate the value of virtual health resource $\rho_{it}^* = \pi_{it}/\varphi_{it}$ for each region and year.
3. Calculate the total virtual health resource ρ_t^* for each year. The ratio of total virtual resources between the individual years will show decrease or increase in the capacity level.
4. Calculate the value of selected inequality measure for the virtual health resource for each year. These values show a decrease or an increase in regional inequality.

Although the values of capacity levels and of inequality measures expressed in the virtual resource cannot be directly interpreted as in the case of original resources, the proposed method has an advantage that it combines all health resources in one dimension.

3 Application

3.1 Data

The Czech Republic is a country with 10.5 million inhabitants (in 2015) that is administratively divided into 14 regions. Prague, the capital, has a status of region. The region of Prague is located in the middle of the territory of the Stredocesky region. We assume that the population of the Stredocesky region use frequently health services in Prague. That is why the data of these two regions, the Prague and Stredocesky regions, were joined together. In the Czech Republic, the health services are reimbursed by public health insurance which should guarantee equal access to health services for the whole population. In the public system, equal distribution of health resources is an important objective of health policy. The data come from the years 2005 and 2015 and were obtained from the Czech Health Statistics Yearbooks [13, 14]. In this study we consider two inputs: the number physicians measured in full-time equivalents and the number of general nurses and midwives measured in full-time equivalents. The regional population that is used as a proxy for health need is the single output.

3.2 Results

The intertemporal production frontier was constructed (Figure 1). First, the technical efficiency scores for all 26 observations were calculated by the input-oriented constant returns-to-scale DEA model (3) with two inputs (the number physicians and the number of nurses) and one output (population). The efficiency scores are presented in Tables 1 and 2. Second, the values of virtual health resource for each region were calculated (Tables 1 and 2). The change in the total capacity level of health resources is measured by the ratio of the total virtual resource in 2015 (12,714,223) to the total virtual resource in 2005 (12,337,383), which is 103.1, so the total virtual resource in 2015 is greater than that in 2015 by 3.1%. The decrease in the number of nurses and midwives was more than compensated by the increase of the number of physicians, thus the Czech population is better-off.

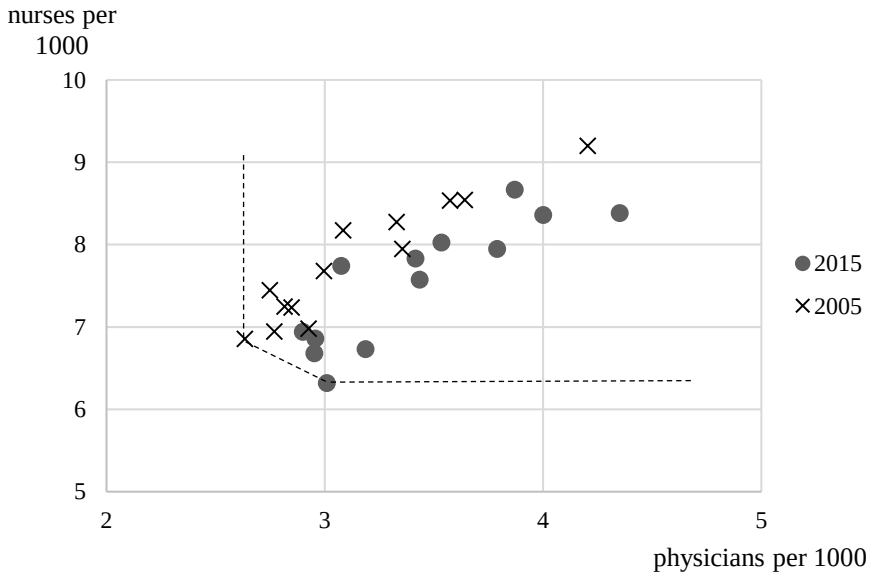


Figure 1 Intertemporal Production Frontier

Region	Population	Number of physicians	Number of nurses	Efficiency Score	Virtual Resource
Prague + Stredocesky	2 326 244	9 778	21 396	0.698	3 330 690
Jihocesky	626 766	1 834	4 374	0.951	658 788
Plzensky	550 371	1 967	4 697	0.778	707 122
Karlovarsky	304 587	939	2 489	0.854	356 774
Ustecky	822 977	2 317	5 964	0.942	873 735
Liberecky	428 268	1 219	3 099	0.939	456 057
Kralovehradecky	547 849	1 824	4 532	0.815	672 345
Pardubicky	505 553	1 400	3 511	0.974	519 128
Vysocina	510 000	1 401	3 798	0.958	532 226
Jihomoravsky	1 130 282	4 116	9 654	0.773	1 463 120
Olomoucky	638 981	2 144	5 078	0.833	766 824
Zlinsky	590 447	1 554	4 047	1.000	590 447
Moravskoslezsky	1 251 767	3 750	9 611	0.888	1 410 126
Czech Republic	**10 234 092**	**34 242**	**82 249**	**x**	**12 337 383**

Table 1 Regional Characteristics, Efficiency Scores, and Virtual Resource, 2005

Region	Population	Number of Physicians	Number of Nurses	Efficiency Score	Virtual Resource
Prague + Stredocesky	2 583 228	11 237	21 654	0.754	3 426 536
Jihocesky	637 292	2 030	4 288	0.941	677 009
Plzensky	575 665	2 181	4 575	0.795	723 897
Karlovarsky	298 506	1 019	2 337	0.835	357 295
Ustecky	823 381	2 386	5 713	0.958	859 291
Liberecky	439 152	1 321	2 775	1.000	439 152
Kralovehradecky	551 270	1 948	4 424	0.812	678 881
Pardubicky	516 247	1 523	3 449	0.974	529 811
Vysocina	509 507	1 566	3 945	0.875	582 402
Jihomoravsky	1 173 563	4 695	9 809	0.756	1 552 184
Olomoucky	635 094	2 458	5 505	0.748	849 164
Zlinsky	584 828	1 729	4 011	0.958	610 470
Moravskoslezsky	1 215 209	4 172	9 203	0.851	1 428 132
Czech Republic	**10 542 942**	**38 268**	**81 688**	**x**	**12 714 223**

Table 2 Regional Characteristics, Efficiency Scores, and Virtual Resource, 2015

Now we will continue with the inequality measurement. For a comparison of usual one-dimensional and proposed multi-dimensional methods, we measure the distribution of physicians and nurses and midwives separately first (Table 3). In 2005, the values of individual Robin Hood Indexes were 7.18% for physicians and 4.52% for nurses and midwives. In 2015, the values of individual Robin Hood Indexes were 6.64% for physicians and 3.96% for nurses and midwives. So we observe an improvement in the geographical distribution of both physicians and nurses and midwives during the years 2005 and 2015. The situation in the case of nurses and midwives is better than for physicians and, in fact, the distribution of nurses and midwives is not so far from equal distribution. The similar results were obtained by the Gini coefficient.

In the proposed inequality measurement method, the Gini coefficient and the Robin Hood Index were applied to the virtual health resource that takes into account that the combinations of two health resources serve regional populations. The regional inequality during the observed decade decreased according to both the Gini coefficient and the Robin Hood Index (Table 3).

Measure/Resource/Year	2005	2015
GC – physicians	0.092	0.084
GC – nurses and midwives	0.058	0.050
GC – virtual resource	0.071	0.056
RHI – physicians	7.18%	6.64%
RHI – nurses and midwives	4.52%	3.96%
RHI – virtual resource	5.53%	4.52%

Table 3 Regional Inequality, 2005 and 2015

4 Conclusion

The method that is able to deal with multiple health resources and measure both the changes of the level of resource capacity and the geographic inequality in the resource distribution was developed. The method is based on the data envelopment analysis methodology that is able to transform multiple health resources into a single virtual resource. In our calculations, the intertemporal production frontier was assumed; however, it is an interesting direction for further research to consider other types of dynamic production frontiers. Regarding the illustrative application to the Czech Republic, we observe an improvement in both the resource capacity and the regional distribution of both physicians and general nurses and midwives between the years 2005 and 2015.

Acknowledgements

The research was supported by the project no. 16-01821S of the Czech Science Foundation.

References

[1] Amin, G. R., & Toloo, M. (2004). A polynomial-time algorithm for finding epsilon in DEA models. *Computers and Operations Research*, 31(5), 803–805. doi: 10.1016/S0305-0548(03)00072-8.
[2] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). Measuring the Inefficiency of Decision Making Units. *European Journal of Operational Research*, 2, 429–444. doi: 10.1016/0377-2217(78)90138-8.
[3] Charnes, A., Cooper, W. W., Lewin, A. Y. & Seiford, L. M. (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Boston: Kluwer Academic Publishers.
[4] Cooper, W. W., Seiford, L.W. & Zhu, J. (2004). *Handbook of Data Envelopment Analysis*. Boston: Kluwer Academic Publishers, 2004.
[5] Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Second Edition. Springer US. doi: 10.1007/978-0-387-45283-8.
[6] De Maio, F. G. (2007). Income inequality measures. *Journal of Epidemiology and Health*, 2007, 61, 849–852. doi: 10.1136/jech.2006.052969.
[7] Dlouhý, M. (2017). A DEA-Based Inequality Measure: Application to Allocation of Health Resources. In P. Prazak (Ed.), *35th International Conference Mathematical Methods in Economics 2017 (MME 2017)*, pp. 125–130. Hradec Králové: University of Hradec Králové.
[8] Dlouhý, M. (2018). Measuring Geographic Inequalities: Dealing with Multiple Health Resources by Data Envelopment Analysis. *Frontiers in Public Health*, 6, article 53. doi: 10.3389/fpubh.2018.00053.
[9] Jablonský, J. & Dlouhý, M. (2015). *Modely hodnocení efektivnosti a alokace zdrojů* [Models of Efficiency Evaluation and Resource Allocation]. Prague: Professional Publishing.
[10] Jílek, J. & Moravová, J. (2007). *Ekonomické a sociální indikátory* [Economic and Social Indicators]. Prague: Futura.
[11] Kawachi, I. & Kennedy, B. P. (1997). The Relationship of Income Inequality to Mortality: Does the Choice of Indicator Matter? *Social Science and Medicine*, 45(7), 1121–1127. doi: 10.1016/S0277-9536(97)00044-0.
[12] Tulkens, H. & Eeckaut, P. V. (1995). Nonparametric efficiency, progress and regress measures for panel data: methodological aspects. *European Journal of Operations Research*, 80(3), 474-499.
[13] ÚZIS ČR - Institute of Health Information and Statistics of the Czech Republic. (2006). *Czech Health Statistics Yearbook 2005*. Prague: ÚZIS ČR.
[14] ÚZIS ČR - Institute of Health Information and Statistics of the Czech Republic. (2016). *Czech Health Statistics Yearbook 2015*. Prague: ÚZIS ČR.

Parameter estimation of regression model with quantized variables

Marek Dvořák[1], Petr Fiala[2]

Abstract. In this paper, we present two novel methods for parameter estimation of binary classification models. Our methods assume that the dependent variable is not binary by nature but a result of quantization of underlying data. The methods are based on multi-stage and moment estimation of the quantized dataset.
The usual classical methods like logistic regression yield rather large variance of the estimated parameters. We have made some adjustments to these classical models in such a way as to reduce this unwanted variance.
The nature of estimated parameters and their variance was examined experimentally on artificial data as well as on real world datasets. These experiments show that the variance of estimated parameters was reduced considerably. Other parameter properties based on these experiments are also described.
A discussion is held to ask and think about several questions and issues regarding our methods as well as some possible theoretical amendments.

Keywords: estimation, quantized variable, logistic regression, variability.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Let us imagine that there is data consisting of bitmap images of portraits of people with labels assigning the gender of depicted person. We might want to create some model which predicts this gender based on image data [9], [7]. Out of several issues that might arise, let us focus on one in particular. Depending on the exact model we chose, out predicted results might be a continues variable between *male* and *female*, which makes a real world sense, since many women may have some male characteristics as well as some men may be more or less feminine.

However our initial assumption was that the labels were strictly binary (or generally quantized). With classical methods, like logistic regression [8], [4], [3], This may result in biased estimations or estimations with high variance. In this paper, we assume that there is an underlying dependent continuous latent variable which is unobserved and that we can only observe the labels which are a quantization of this latent variable [11], [2].

We first take a look at the formal definition of the problem. Then we introduce a specialized estimator for single variable, followed by generalized estimator for two or more variables. Finally, we take a look at some properties of this estimator, namely the variance of estimations, which we show experimentally.

2 Problem definition

Let us define a simple linear model in equation 1. Note that $X_{i0} = 1$ for all i and as such b_0 can be interpreted as an intersect.

$$Y_i = \sum_{j=0}^{m} b_j X_{ij} + e_i, \tag{1}$$

where $i \in 1, \dots n$ denotes the observation, $j \in 1, \dots m$ denotes the explanatory variables and e_i signifies a normal distributed error term with zero mean and non-negative variance.

$$e \sim \mathcal{N}\left(0, \sigma^2\right) \tag{2}$$

We can then define a quantization function in equation 3:

[1]University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 1938/4, Prague, Czech Republic, marek.dvorak@vse.cz
[2]University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 1938/4, Prague, Czech Republic, pfiala@vse.cz

$$Y_i' = \begin{cases} -1 & \text{if } Y_i \leq 0 \\ 1 & \text{if } Y_i > 0 \end{cases} \tag{3}$$

The problem is then to estimate $b_0, \dots b_m$ given the observations pairs (Y_i', X_{ij}) for all i and j. If we want to estimate the model using binomial logistic regression in the form of equation 4:

$$\log\left(\frac{p_i}{1-p_i}\right) = \sum_{j=0}^{m} \beta_j X_{ij}, \tag{4}$$

where p_i is the probability of observation Y_i' being 1 given X_i, we can use the maximum likelihood function 5:

$$l\left(\beta\right) = \sum_{i=1}^{n} \left[\left(\frac{1}{1+e^{-\sum_{j=0}^{m} b_j X_{ij}}} \right)^{Y_i} \left(1 - \frac{1}{1+e^{-\sum_{j=0}^{m} b_j X_{ij}}} \right)^{1-Y_i} \right]. \tag{5}$$

We can then maximize this maximum likelihood function given the parameters b_j. Logistic regression yields a parameter estimation whose properties have been thoroughly studied [5],[1], [10], [6].

In this paper we would like to introduce another method for the parameter estimation and experimentally look at its properties.

3 Parameter estimation

The whole estimator is build upon the idea of fitting a hyperplane to two points in hyperspace, where the two points represent mean of positive and negative examples. This works fine in two dimensional space, since a line is defined using two points, but causes problem in higher dimensional space, because a plane is defined using three points. Because of this, we introduce a restriction that the fitted hyperplane normal is coplanar with another hyperplane also defined by those two points and which is also perpendicular to yet another plane defined by two vectors - representing axes of explanatory variables. This mouthful definition is to introduce a simple concept which in less rigorous terms could be explained that the fitted hyperplane should 'lay as flat as possible'. The first variation worth considering is when there is one external explanatory variable X_{i1} with two parameters b_0, b_1. When there is only a single variable b_0 - the intercept - the b_0 parameter estimation would reduce to $[b_0] = \frac{0}{0}$ as can be shown from the generalized model equations 9 to 19.

3.1 Single explanatory variable

When we have a single explanatory variable, we can construct a specialized estimator like this:

$$EX^+ = \frac{1}{n^+} \sum_{i^+=1}^{n^+} X_{i^+,1} \tag{6}$$

$$EX^- = \frac{1}{n^-} \sum_{i^-=1}^{n^-} X_{i^-,1} \tag{7}$$

$$[b_0, b_1] = \left[\frac{EX^- + EX^+}{EX^- - EX^+}, \frac{2}{EX^+ - EX^-} \right], \tag{8}$$

where n^+ denotes the observations with a positive label and n^- denotes the observations with a negative label.

It is worth noting that estimations b_0 and b_1 are not fully identified as there is one degree of freedom. As it turns out, if we introduce arbitrary scale index m, the estimations 8 can be corrected as $[mb_0, mb_1]$.

3.2 Two explanatory variables

With two explanatory variables, we can generalize equations 6 to 8 using matrix notation with equations 9 through 19.

$$EX_1^+ = \frac{1}{n^+} \sum_{i^+=1}^{n^+} X_{i^+,1} \tag{9}$$

$$EX_1^- = \frac{1}{n^-} \sum_{i^-=1}^{n^-} X_{i^-,1} \tag{10}$$

$$EX_2^+ = \frac{1}{n^+} \sum_{i^+=1}^{n^+} X_{i^+,2} \tag{11}$$

$$EX_2^- = \frac{1}{n^-} \sum_{i^-=1}^{n^-} X_{i^-,2} \tag{12}$$

$$A = \left[EX_1^-, EX_2^-, -1\right] \tag{13}$$

$$B = \left[EX_1^+, EX_2^+, 1\right] \tag{14}$$

$$C = \frac{A+B}{2} \tag{15}$$

$$Z = [0, 0, 1] \tag{16}$$

$$N = (B-A)\left((B-A)\,Z\right) - Z\,(B-A)^2 \tag{17}$$

$$c = NC \tag{18}$$

$$[b_0, b_1, b_2] = \left[\frac{c}{N_3}, \frac{N_1}{N_3}, \frac{N_2}{N_3}\right]. \tag{19}$$

Again, these three estimated coefficients are not fully identified, but there is still only one degree of freedom. As was the case in the section 3.1, we can introduce a scale index m which fills the remaining degree of freedom with the full estimator being $[mb_0, mb_1, mb_2]$.

In fact, if we take the estimator in this section and lower the number of explanatory variables, we can reduce the whole system equations 9 to 19 to the system of equations 6 to 8. And analogously, we can quite easily generalize the estimator for arbitrary natural number of explanatory variables.

4 Experiments on artificial data

To examine the properties of estimated parameters, we have designed a numerical experiment on artificial data.

The experiment started with creating a 1000 observations of artificial variables $X_1 \sim \mathcal{N}(0, 2)$, $X_2 \sim \mathcal{N}(0, 2)$ and $e \sim \mathcal{N}(0, 1)$. Then an explanatory variable were created using equations 20 and 21 with parameters set to $[b_0, b_1, b_2] = [1, 4, 4]$.

$$Y_i = 1 + 4X_{i1} + 4X_{i2} + e_i \tag{20}$$

$$Y_i' = \begin{cases} -1 & \text{if } Y_i \leq 0 \\ 1 & \text{if } Y_i > 0 \end{cases} \tag{21}$$

Next, the parameters of these three models were estimated.

1. Least squares linear regression without quantized explanatory variables
2. Binomial logistic regression with quantized variables
3. Our estimator with quantized variables

These models were estimated each 10000 times to get statistics for each estimated parameter.

Graphical results of this simulation are shown in figures 1, 2 and 3. The figure 1 shows a histogram of 10000 estimated b_0 parameters for both examined models. The ground-truth value of b_0 was 1. The same applies to 2 and 3 except here, the ground truth for b_1 and b_2 was 4. We can see that the parameter estimations using classical logistic regression (the red histogram) have much higher variance than estimation using the method presented in this paper (the blue histogram).

We can also look at the mean and standard deviation in table 1. The table shows that the method described in this paper provides parameter estimation with lower variance compared to the logistic regression.

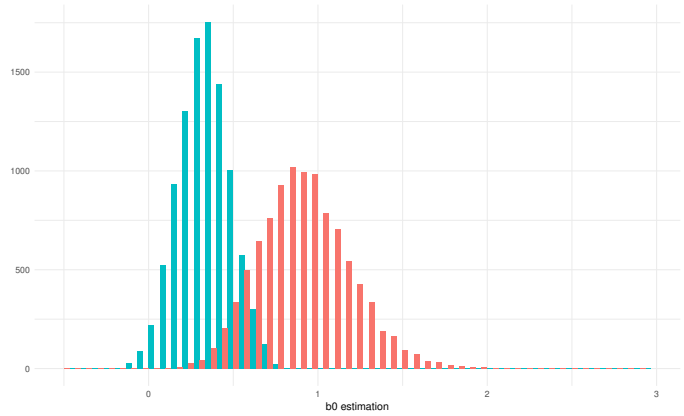


Figure 1 Histograms of the estimated b_0 values. The red (rightmost) is the histogram of logistic regression, the blue (leftmost) is the histogram of out method.

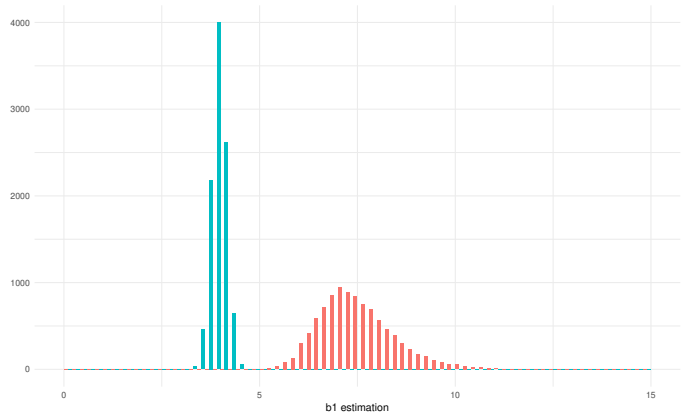


Figure 2 Histograms of the estimated b_1 values. The red (rightmost) is the histogram of logistic regression, the blue (leftmost) is the histogram of out method.

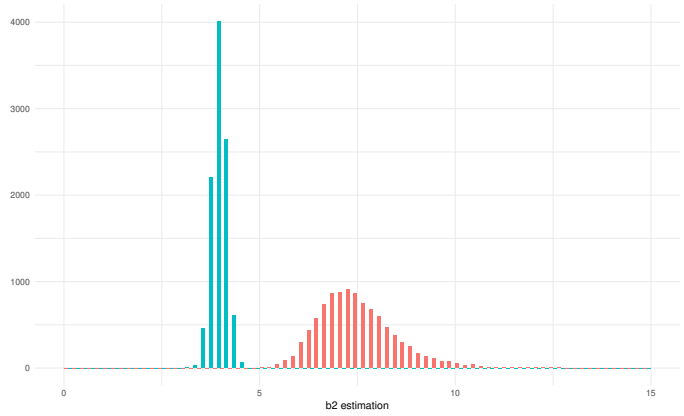


Figure 3 Histograms of the estimated b_2 values. The red (rightmost) is the histogram of logistic regression, the blue (leftmost) is the histogram of out method.

	mean	standard deviation
b_0 lm	1.000	0.0377
b_1 lm	3.999	0.0270
b_2 lm	4.000	0.0268
b_0 logistic	0.941	0.2740
b_1 logistic	7.529	0.9905
b_2 logistic	7.532	0.9944
b_0 ours	0.305	**0.1528**
b_1 ours	3.915	**0.1898**
b_2 ours	3.918	**0.1905**

Table 1 Table showing the properties of three estimations.

5 Conclusion

In this paper, we have presented an estimator for regression with quantized dependent variable as opposed to more traditional linear regression. We have also used numerical simulation of artificial data to determine the properties of this model. We have experimentally shown that this method can lead to estimations with lower variance then logistic regression.

Further work might include calculating the exact estimator properties analytically as well as determining its behaviour on real data or with large number of explanatory variables.

Acknowledgements

Supported by the grant No. IGA F4/58/2017 of the Internal Grant Agency, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Allison, P. D. (2012). Logistic regression using SAS: *Theory and application*. SAS Institute.

[2] Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, 45(3), 289-308.

[3] Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, 35(5-6), 352-359.

[4] Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association*, 70(352), 892-898.

[5] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

[6] Lemeshow, S., & Hosmer Jr, D. W. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *American journal of epidemiology, 115*(1), 92-106.

[7] Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 34-42).

[8] Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.

[9] Moghaddam, B., & Yang, M. H. (2000). Gender classification with support vector machines. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* (pp. 306-311). IEEE.

[10] Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology, 49*(12), 1373-1379.

[11] Wold, H. (1975). Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. *Journal of Applied Probability*, 12(S1), 117-142.

Simulation of investment project internal rate of return

StanislavaDvořáková[1], Petr Jiříček[2]

Abstract. The paper will deal with variants of finding the internal rates of return of investment projects with a subsidy from public sources. The simulations will correspond to gradual changes of the subsidy amounts and the changes of cash flows in individual project phases. The simulations will result in displaying functional dependencies of internal rates of return on multi-parametric project changes. Then the extreme values and limits of these functional dependencies will be determined.

Keywords: simulation of internal rate on return, investment projects, extreme values.

JEL classification: C20, H43
AMS classification: 65H04

1 Introduction

For the evaluation of commercial investment projects, two main methods are used: the net present value method and the internal rate of return method. The internal rate of return method is based on finding such a real root of the NPV function from the proposed project parameters that corresponds to the condition of indifferent evaluation of the public project efficiency from the investors perspective. The recommendation for evaluating projects by the commercial method of IRR reflects the fact that the IRR method has, in comparison with the NPV method, several significant disadvantages. These relate namely to the so-called non-conventional cash flows generated by some projects, when more than one change of cash flow polarity occur. Bussey and Eschenbach (1992) define the conventional type of investment as an investment that contains one or more negative cash outflows, followed by one or more positive cash inflows. A non-conventional investment is defined as an investment that intersperses the positive and negative cash flows. Problems of the IRR method were completely dealt with in the past e.g. by Teichroew et al. (1965) in compliance with solving the polynomial by the so-called Descartes rule. Hazen (2003) has solved the problem of multiple real IRRs by linking the present value of the outstanding capital expenditure of a project with the difference between any IRRs and the cost of capital. The topic of using IRR and NPV criteria has been summarised by Magni (2011); with a view of the focus of this paper, we can mention especially these points:

1. in case of multiple roots of the NPV function in the area of the positive halfplane corresponding to two or more values of a given projects IRR, multiple realvalued IRRs may arise;
2. the NPV function has complex-valued roots and so complex-valued IRRs may arise;
3. the decision criterion in case of the IRR is not compatible with the decision criterion of the NPV method in the relation to the standpoint to reject or accept the project;
4. the IRR method does not produce the same ranking of mutually exclusive projects as the NPV method and so it is not in conformity with the goal of maximizing the company market value or social welfare in case of public projects.

Mishan and Quah (2007) address the topic of multiple-valued roots of the NPV function for using the IRR criteria to evaluate public projects by means of cost-benefit analysis. Pearce et al. (2006) was concerned with using the IRR method for evaluating development public projects (called the economic rate of return, ERR) by means of cost-benefit analysis based on theoretical principles of the utility theory and the Kaldor-Hicks compensation test. The present methodology of the European Union results namely from works of Florio (2006) and Florio and Vignetti (2013). The resulting evaluation of social efficiency of public projects is carried out based on discounting economic cash flows by means of the ENPV (an NPV alternative used for public projects) and ERR (the European Commission, 2015).

2 Objective and methods

The simulations will result in displaying functional dependencies of internal rates of return on multi-parametric project changes. Then the extreme values and limits of these functional dependencies will be determined. The Net Present Value method in case of public projects ($ENPV$) is expressed by the investment curve

[1]College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 586 01 Jihlava, Czech Republic, stanislava.dvorakova@vspj.cz
[2]College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, jiricek@vspj.cz

$$ENPV = \sum_{t=0}^{n} \frac{CF_t}{(1+k)^t}, \tag{1}$$

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.

Stream of CFs – the sequence of cash flows generated by the public project in the form of $CF = [CF_1, CF_2, CF_3, CF_4, CF_5, CF_6]$, in the model in the basic form $CF = [-6, 1, 1, 1, 1, 1, 1]$, which corresponds to realistically predictable flows of proposed public projects.

Economic internal rate of return of the project (ERR) – the root of the investment curve (1) of the project, i.e

$$ENPV = \sum_{t=0}^{n} \frac{CF_t}{(1+ERR)^t} = 0. \tag{2}$$

The presented paper aims at simulating internal rate of return for public projects (ERR) by means of modelling the root function.

1. The root function $CF_j = \varphi(ERR)$, $0 \leq j \leq n$ – the function of roots for the function of two variables of $ENPV(ERR, CF_j) = 0$, i.e.

$$CF_j = \varphi(ERR) = -\sum_{t=0}^{j-1} \frac{CF_t}{(1+ERR)^{t-j}} - \sum_{t=j+1}^{n} \frac{CF_t}{(1+ERR)^{t-j}}. \tag{3}$$

2. Therefore it is a curve corresponding to the intersection of the surface of $ENPV(k, CF_j)$ and the plane of $ENPV = 0$.
3. Conventional projects – they are characterized by only one change of their polarity in the sequence of the stream of cash flows generated within the project, e.g. $[-, - - + + ++]$).
4. Non-conventional projects – they are characteristic by more than one change of project generated cash flows in the sequence of the CF stream (e.g. $[-, - + + + +-]$).

In the simulation, we seek the dependence of the curve course and root position on the amount of the subsidy from public sources, manifesting in the change of CF_0 for the cases of:

- Conventional project with a crisis in the building phase $CF = [CF_0, - + + + ++]$,
- Non-conventional project with a crisis in the operational phase $CF = [CF_0, + + - + ++]$,
- Non-conventional project with a crisis $CF = [CF_0, + + + + +-]$.

We change the parameters of the stream of cash flow according to selected basic scenarios for individual levels of subsidy from public sources of the European Union and we display the course of the root function for individual variants of projects. We thus seek the number and position of the ENPV function roots for individual levels of subsidy in reaction to the change of the critical parameter of individual scenarios, its extreme values and limits, i.e. in reaction to:

- the change of cash flow in the public project building phase (ΔCF_1) – e.g. caused by critical delayed construction of the investment (scenarios with 100%, 70%, 30%, 0% subsidy from the European Union); this dependence is expressed by the function $CF_1 = \varphi(ERR)$,
- changes of cash flow in a selected year of the operational phase of the project (ΔCF_3) – e.g. caused by an economic crisis (scenarios with 100%, 70%, 30% , 0% subsidy from the European Union); this dependence is expressed by the function $CF_3 = \varphi(ERR)$,
- changes of cash flow in the liquidation phase of the project (ΔCF_6) – e.g. caused by high environmental costs (scenarios with 100%, 70%, 30%, 0% subsidy from the European Union); this dependence is expressed by the function $CF_6 = \varphi(ERR)$.

3 Results and discussions

To get a general overview, we present a projection of a set of four root functions for selected levels of subsidy of a public project (100%, 70%, 30%, 0%) in a summary graph. In the graph, we can see the influence of the subsidy level on the course of root functions for individual scenarios (see Fig. 1, 2, 3).

The root curves are not defined for $ERR = -1$ (which results from relation (3)). We will judge the course of the curves only for $ERR > -1$ because negative rates of return are not economically significant in the investment project area.

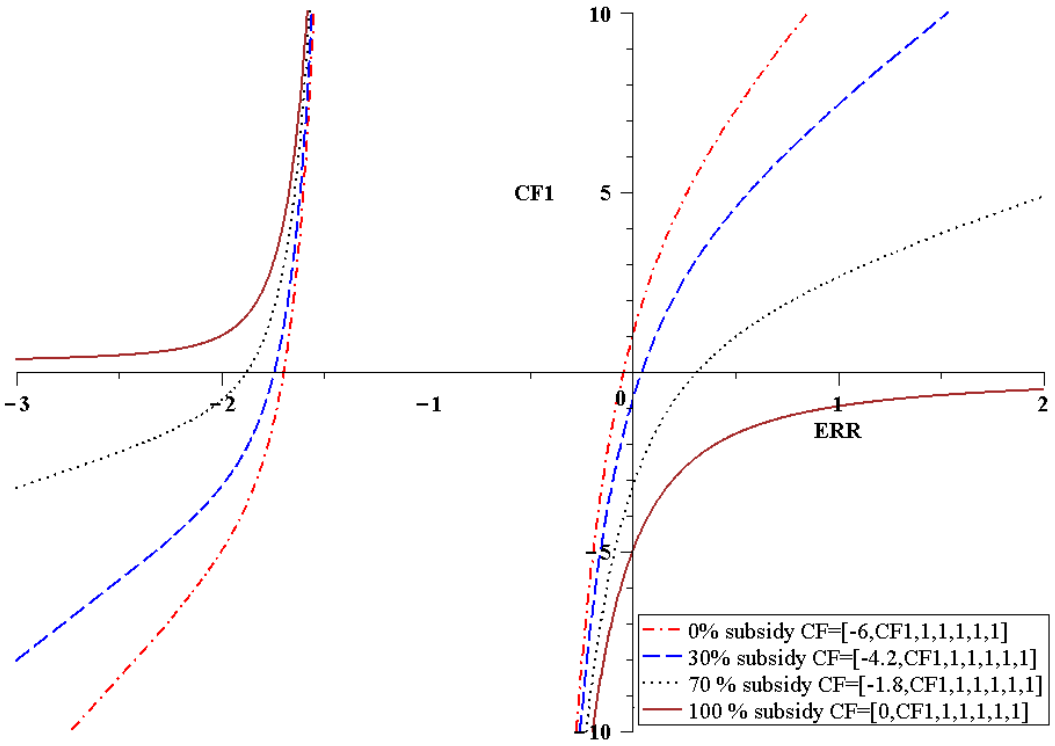


Figure 1 Quadruplet of functions $CF_1 = \varphi(ERR)$ for various parameters of subsidy amount (crisis in a building phase of public investment project)

All the functions in Fig. 1 are not defined for the value $ERR = -1$ and thus their limit here is vertical. The functions intersect the y-axis (i.e. $ERR = 0$) at the value of $CF_1 - \sum_{t=0}^{n} CF_t$, i.e. at values 1, -0.8, -3.2 and -5. All the root functions have the limit of ∞ for $ERR \to \infty$ except for the root function for a 100% subsidy. This curve has the limit equal to zero for $ERR \to \infty$. In this simulation, the root functions do not have extremes.

All the functions in Fig. 2 are also not defined at $ERR = -1$, therefore they have a vertical limit here. The functions intersect the y-axis (i.e. $ERR = 0$) at the point $CF_3 - \sum_{t=0}^{n} CF_t$, i.e. at values 1, -0.8, -3.2, and -5. All the root functions have the limit of ∞ for $ERR \to \infty$ except for the curve of the 100% subsidy. This curve has a limit equal to $-\infty$ for $ERR \to \infty$ and has a local maximum in the point $ERR = 0.19$ and $CF_3 = -4.746$. For the values of $CF3 > -4.2$ (root function for 100% subsidy), the $ENPV$ function (1) has two roots, which means that there are two ERRs and thus the efficiency of the project cannot be unambiguously determined by this method.

Unlike the previous cases, the functions in Fig. 3 are defined for all ERR values (this results from relation (3)). For $ERR = -1$, CF_6 is equal to zero. The functions intersect the y-axis (i.e. $ERR = 0$) at the value $CF_6 - \sum_{t=0}^{n} CF_t$, i.e. at values 1, -0.8, -3.2, and -5. All the root functions have the limit of ∞ for $ERR \to \infty$ except the curve for the 100% subsidy. This curve has the limit equal to $-\infty$ for $ERR \to \infty$. All the curves except the curve of the 100% subsidy have a local minimum (curve for 70% subsidy in the point $CF_6 = -3.592$ and $ERR = 0.16$), i.e. $ENPV$ function (1) has two ERR roots.

4 Conclusion

The IRR method (according to the EU methodology the ERR) is the second most preferred method of public investment evaluation; the reason is well-known disadvantages of this method in case of non-conventional projects, mentioned in this article. Yet, some theoretical and practical concepts prefer, as stated in the article, the IRR (ERR) method. The paper solves non-conventional public projects when negative cash flows arise due to crisis situations in operational and liquidation phase of a public project. The paper also presents a comparison with public projects of conventional type (in case of the crisis in a building phase).

Multi-parametrical analysis of a project is carried out on the grounds of displaying the $ENPV$ function which

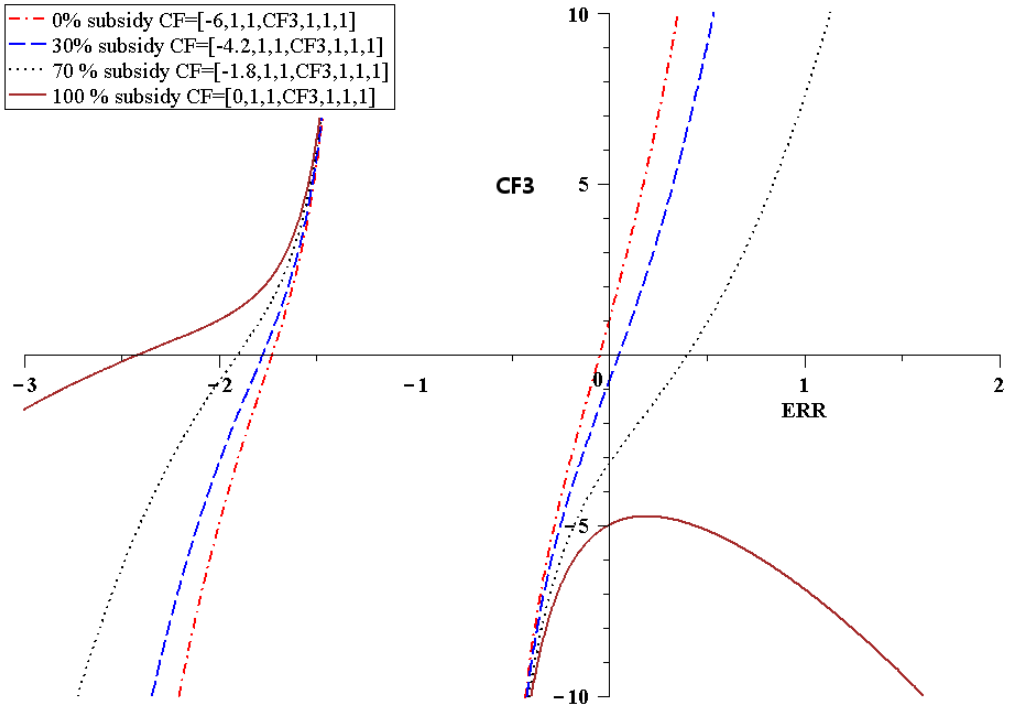


Figure 2 Quadruplet of functions $CF_3 = \varphi(ERR)$ for various parameters of subsidy amount (crisis in a operational phase of public investment project)

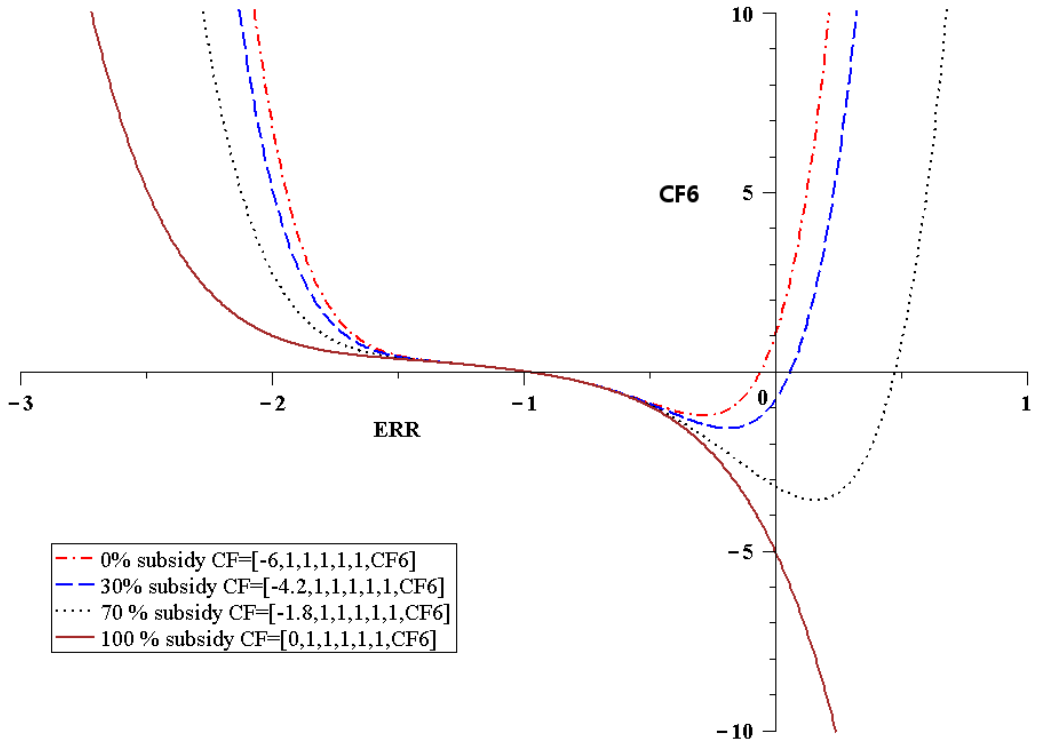


Figure 3 Quadruplet of functions $CF_6 = \varphi(ERR)$for various parameters of subsidy amount (crisis in a liquidation phase of public investment project)

models a public project given by a stream of cash flows in the basic version $CF = [-6, 1, 1, 1, 1, 1, 1]$ in 3D projection, where the influence of change of project parameters on the position and curving of the ENPV surface can be seen.

The function of the ENPV roots gives information about the position, number and value of the ERR in individual model scenarios. This value can be read directly from the graph of the root function $CF_j = \varphi(ERR)$, $0 \leq j \leq n$ subject to the change of individual model parameters (i.e. cash flows of the real project).

From the results of the analysis, it can be demonstrated that when parameters (CF_j) change there is a field which allows us to apply the ERR even for non-conventional projects, i.e. the $ENPV$ function gives exactly one real root > 0. Such projects are characterized by relative stability. The occurrence of more positive real roots, as shown by the results of the analysis of functions of the ENPV roots in the model, would probably be caused by much higher instability of projects, which should not be the rule with public projects, regarding positive externalities. The conditions for the rise of such instabilities can be determined on the basis of the proposed model generally by means of the function $CF_j = \varphi(ERR)$ by further analysis.

The rise of such an instability can be shown, as has already been mentioned, e.g. from the course of the root function $CF_6 = \varphi(ERR)$ for non-conventional projects with crisis liquidation phase with 70% subsidy (Fig. 3). Here it is apparent that in a certain defined field of the CF_6 parameter the root function has a local minimum, so for one value of CF_6 there obviously are 2 values of ERR.

The model based on applying a rational fractional function for the description of a public project thus allows us to carry out consequent analysis of the function $CF_j = \varphi(ERR)$, an analysis of different variants of public projects as well, which are not solved in the article. These are e.g. non-conventional projects with multiple changes of cash flow polarity, which are characterized by more than two real roots.

The use of the EER method generally depends on the preference for the utility of either the investor or the donor of the project. When comparing multiple projects, the ERR method (which prefers higher rate of return within the projects return period) can provide a different order of selection of the optimal project alternative from the ENPV method, which prefers the absolute earnings during the required lifetime of the project.

The optimal amount of support with the specified parameters of the project (lifetime period, amount of investment costs, cash flow from the project) can be determined by means of sensitivity analysis of the project based on finding the projects threshold value.

Acknowledgements

Supported by the grant No. IGS 1170/04/189 of the College of Polytechnics Jihlava.

References

[1] Dvořáková, S. & Jiříček, P. (2013): Modelling Financial Flows of Development Projects Subsidized from European Funds. In H. Vojáčková (Eds.), *Proceedings of the 31st International Conference on Mathematical Methods in Economics* (pp. 119–128). Jihlava: Vysoká škola polytechnická.

[2] European Commission (2015). *Guide to Cost-Benefit Analysis of Investment Project*. Brusel: DG Regional and Urban Policy.

[3] Florio, M. & Vignetti, S. (2006). Cost-benefit analysis of infrastructure projects in an enlarged European Union: Returns and Incentives. *Economic change and restructuring 38*, 179–210.

[4] Florio, M. & Vignetti, S. (2013). *The use of ex-post Cost-Benefit Analysis to assess the longterm effects of Major Infrastructure Projects*, available at http://www.csilmilano.com/docs/WP2013_02.pdf.

[5] Hazen, G. B. (2003). A new perspective on multiple internal rates on return. *The Engineering Economist 48(1)*, 31–35.

[6] Magni, C. A. (2011). Average Internal Rate of Return and investment decisions: a new perspective. *The Engineering Economist 55(2)*, 150–180.

[7] Novotná, J. & Trch, M. (2000). *Algebra a teoretická aritmetika. Polynomická algebra*. Praha: Univerzita Karlova.

[8] Pearce, D. & Atkinson, G. & Mourato, S. (2006): *Cost-Benefit Analysis and the Environment*. Available at http://www.lne.be/themas/beleid/milieueconomie/downloadbare-bestanden/ ME11_cost-benefit%20analysis%20and%20the%20environment%20oeso.pdf.

[9] Šisler, M. & Andrys, J. (1996). *O řešení algebraických rovnic*. Praha: Mladá fronta.

[10] Teichroew, D. & Robichek, A. A. & Montalbano, M. (1965). An analysis of criteria for investment and financing decisions under certainty. *Management Science 12(3)*, 151–179.

Modeling of supply network coordination

Petr Fiala[1], Renata Majovská[2]

Abstract. Supply network is a decentralized system composed from layers of potential suppliers, producers, distributors, retailers and customers etc., where agents are interconnected by material, financial and information flows. These agents have different interests and the whole network behaves ineffective. The expanding importance of supply network coordination presents a challenge to focus more attention on supply network modeling. The game theory approach seems to be a basis for other approaches. Supply network behavior is influenced by a combination of competitive and cooperative relationships between agents. Biform games combine non-cooperative and cooperative approaches of game theory and are promising for modelling behavior of the agents. It may form a global view of the coordination problem. Other approaches to coordination such as auctions and contracts are possible to model as specific game theory models.

Keywords: supply network, coordination, game theory, auctions, contracts.

JEL Classification: C44, L14, D44, C61
AMS Classification: 90C15

1 Introduction

Supply network is a decentralized system composed from layers of potential suppliers, producers, distributors, retailers and customers etc., where agents are interconnected by material, financial and information flows. A supply network is the collection of steps that a system takes to transform raw components into the final product. There is much inefficiency in supply network behaviour. When one or more agents of the supply network try to optimize their own profits, system performance may be hurt.

Supply network management has generated a substantial amount of interest both by managers and by researchers. There are many concepts and strategies applied in designing and managing supply networks (see [11], [14]). The expanding importance of supply network coordination presents a research challenge to focus more attention on supply network modelling (see [15], [16] [18]).

Recently, considerable attention of researchers is drown to provide some incentives to adjust the relationship of supply network agents to coordinate the supply network, i.e., the total profit of the decentralized supply network is equal to that achieved under a centralized system. The paper compares coordination mechanisms based on models of game theory, contracts and auctions.

The rest of the paper is organized as follows. Section 2 presents coordination problem in supply networks. Section 3 summarizes some of the basic non-cooperative and cooperative concepts of the game theory, including biform games. Supply network contracts have drawn much attention from the researchers. A contract between members of a supply network with price-dependent stochastic demand is analyzed in Section 4. A complex trading model based on using of iterative process for combinatorial double auctions between members of a supply network is presented in Section 5. Section 6 presents conclusions.

2 Coordination problem

The most important part of managing phase is the coordination of individual activities to be optimal in terms of the whole system. Supply networks are decentralized systems. A centralized system can be taken as a benchmark. The question is: How to coordinate the decentralized supply network to be efficient as the centralized one? The seller rarely has complete information about buyer's cost structure. However, the quantity the buyer will purchase and therefore seller's profit depend on that cost structure. Somehow, the seller will have to take this information asymmetry into account. The numbers of sellers and buyers are denoted by m, n, respectively. The symbol S_i

[1] University of Economics, Prague, department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech Republic, pfiala@vse.cz.
[2] University of Finance and Administration, Prague, department of Computer Science and Mathematics, Estonska 500, 101 00 Praha 10, Czech Republic, renata.majovska@mail.vsfs.cz.

represents *i*-th seller while the symbol B_j represents *j*-th buyer. The seller-buyer relations in supply network can be taken as decentralized or centralized (see Fig. 1).

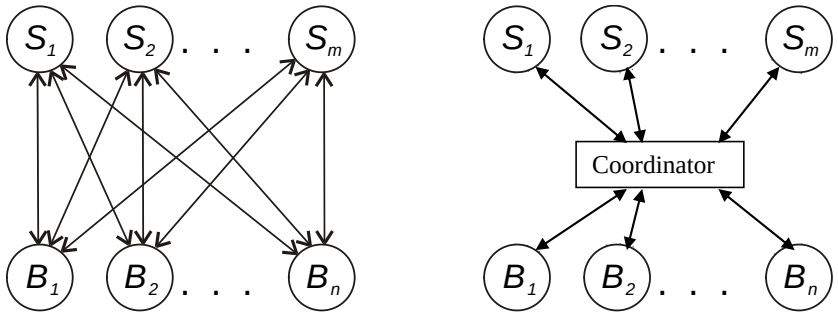


Figure 1 Decentralized (a) and centralized (b) seller-buyer relations

Most supply networks are composed of independent agents with individual preferences. It is expected that no single agent has the power to optimize the supply network. Each agent will attempt to optimize his own preference, knowing that all of the other agents will do the same. This competitive behavior does not lead the agents to choose polices that optimize overall supply network performance due to supply network externalities. The typical solution is for the agents to agree to a set of transfer payments that modifies their incentives, and hence modifies their behavior. Coordination between sellers and buyers can be provided through information sharing (schematically see Fig. 2).

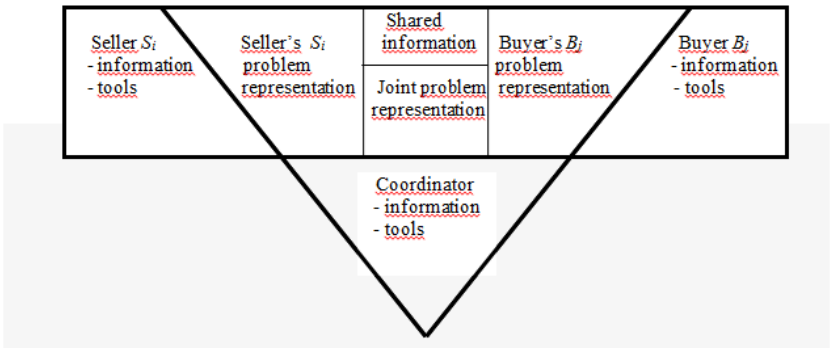


Figure 2 Coordination through information sharing

A seller S_i and a buyer B_j have information and analytical tools for their problem representations. A coordinator helps by information sharing and by formulation of a joint problem representation (see [6]). We compare some coordination mechanisms based on: games, contracts and auctions.

3 Game theory models

This section summarizes some of the basic non-cooperative (see [4]) and cooperative concepts (see [12]) of the game theory that are applied in coordination of supply networks. The non-cooperative theory of games is strategy oriented; it studies what one may expect the players to do. The cooperative theory is a "micro" approach in that it focuses on precise descriptions of what happens.

An n-player non-cooperative game in the normal form is a collection, where N is a set of n players; X_i, $i=1,2,...,n$, is a set of strategies for player i; $\pi_i(x_1,x_2,\ldots,x_n)$, $i=1,2,...,n$, is a pay-off function for player i, defined on a Cartesian product of n sets X_i, $i=1,2,...,n$.

Decisions of other players than player i are summarized by a vector

$$x_{-i}=(x_1,...,x_{i-1},x_{i+1},...,x_n). \tag{1}$$

A vector of decisions $\left(x_1^0, x_2^0, ..., x_n^0\right)$ is the Nash equilibrium of the game if

$$x_i^0\left(x_{-i}^0\right) = \text{argmax}_{x_i}\pi_i\left(x_i, x_{-i}\right)\forall i = 1,2,\ldots,n. \tag{2}$$

The 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. When modelling 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 of n players N is such function $\nu(S)$ that is defined for all subsets $S \subseteq N$ (i.e. for all coalition) and assigns a value $\nu(S)$ with following characteristics:

$$\nu(\varnothing) = 0,\ \nu(S_1 \cup S_2) \geq \nu(S_1) + \nu(S_2), \tag{3}$$

where S_1, S_2 are disjoint subsets of the set N.

The pair (N, ν) is called a cooperative game of n players in the characteristic function form. A particular allocation policy, introduced by Shapley [13], has been shown to possess the best properties in terms of balance and fairness. So called Shapley vector is defined as

$$h = (h_1, h_2, \ldots, h_n), \tag{4}$$

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:

$$\nu(S) - \nu(S - \{i\}). \tag{5}$$

A complicating factor is that with the increasing number of n players is rapidly increasing number of coalitions and complexity of their production. Shapley value for the i-th player is calculated as a weighted sum of marginal contributions according to the formula:

$$h_i = \sum_S \left\{ \frac{(|S|-1)!(n-|S|)!}{n!} \cdot \left[\nu(S) - \nu(S - \{i\})\right] \right\}, \tag{6}$$

where the number of the coalition members is marked by symbol $|S|$ and the summation runs over all coalition $i \in S$.

The ongoing actions in the supply network are a mix of cooperative and non-cooperative behaviour of the participants. A biform game is a combination of non-cooperative and cooperative games, introduced by Brandenburger and Stuart [2]. It 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. The biform game approach can be used for modelling general buyer-seller relationships in supply networks (see [7]). First, sellers make initial proposals and take decisions. This stage is analysed using a non-cooperative game theory approach. Then, sellers negotiate with buyers. 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 seller's share of the total surplus is the product of its added value and its relative negotiation power.

Confidence indices $0 \leq \alpha^i \leq 1$, for all i = 1, 2, ..., n, are introduced. The indices show players' anticipation of the pay-off they will receive in the cooperative stage, i.e. the proportion of the difference between the maximum and minimum core allocation achievable for players. Confidence indices of the players provide the link between the non-cooperative and cooperative stages of the biform game.

4 Contract models

Supply network contracts have drawn much attention from the researchers (see [3]). A combined problem of supply network coordination with price-dependent stochastic demand is analysed (see [8]). We define the following quantities: q buyer's total order quantity; c seller's unit production cost; p retail price. The setting can be characterized as a newsvendor problem. We will analyse the multiplicative form of price-dependent stochastic demand

$$D(p,u) = y(p)u\,, \tag{7}$$

a function of p and u, where u is a random variable independent of p and $y(p)$ is continuous, nonnegative, twice differentiable function. The expectation of D is specified by a function $y(p)$ for any given price p:

$$E\left[D(p,u)\right] = y(p)\,. \tag{8}$$

The expected profit for centralized solution for any output level q and price p is:

$$\begin{aligned}\pi(p,q) &= E\left\{p\left[\min\left(q, D(p,u)\right)\right] - cq\right\} = E\left\{(p-c)q - p\max\left(0, q - D(p,u)\right)\right\} = \\ &= (p-c)q - py(p)\int_0^{\frac{q}{y(p)}} F(u)\,du\end{aligned} \tag{9}$$

The objective is to choose $\left(p^0, q^0\right)$ to maximize the expected profit $\pi(p,q)$.

By fixing price p the problem reduces to standard newsvendor problem without pricing and the optimal level of production

$$q^0 = y(p)F^{-1}\left(\frac{p-c}{p}\right). \tag{10}$$

By substituting it into the expected profit

$$\pi(p) = y(p)\left[(p-c)F^{-1}\left(\frac{p-c}{p}\right) - p\int_0^{F^{-1}\left(\frac{p-c}{p}\right)} F(u)\,du\right]. \tag{11}$$

The problem is now with only one decision variable p and the optimal price p^0 can be obtained by solving

$$\frac{d\pi(p)}{dp} = 0\,. \tag{12}$$

The assumptions of the existence and uniqueness of the optimal solution $\left(p^0, q^0\right)$ are concavity of deterministic part of demand function $y(p)$ and IGFR property of stochastic part of the demand function u.

The proposed contract for coordination of the decentralized supply network is a specific buy-buck contract. The wholesale price w and the buy-buck price b are specified:

$$w = \lambda(p-c) + c\,,\ b = \lambda p\,, \text{ where } 0 \le \lambda \le 1. \tag{13}$$

By the setting of the prices w and b the buyer's profit and the seller's profit for any chosen output level q and price p are

$$\pi_B = E\left\{p\left[\min\left(q, D(p,u)\right)\right] - wq + b\max\left(0, q - D(p,u)\right)\right\} = (1-\lambda)\pi\,, \tag{14}$$

$$\pi_S = E\left\{(w-c)q - b\max\left(0, q - D(p,u)\right)\right\} = \lambda\pi\,. \tag{15}$$

From previous expressions of the buyer's profit and the seller's profit, it is clear that the buyer and the seller solve the same problem as the centralized supply network and the sum of the buyer's profit and the seller's profit is equal to the profit of the centralized supply network. The parameter λ characterizes a splitting of the total profit between the buyer and the seller.

5 Auction models

Combinatorial auctions are promising for supply network analyses (see [5], [9]). We propose a complex coordination model based on using of iterative process for combinatorial double auctions (see [10]). Let us suppose that m potential sellers $S_1, S_2, \ldots S_m$ offer a set R of r items, $j = 1,2,...,r$, to n potential buyers $B_1, B_2, \ldots, B_n$.

A bid made by seller S_h, $h = 1,2,\ldots,m$, is defined as $b_h = \{C, c_h(C)\}$, a bid made by buyer B_i, $i = 1,2,...,n$, is defined as $b_i = \{C, p_i(C)\}$, where $C \subseteq R$ is a combination of items, $c_h(C)$ is the offered price by seller S_h for the combination of items C, $p_i(C)$ is the offered price by buyer B_i for the combination of items C. Binary variables are introduced for model formulation:

$x_i(C)$ is a binary variable specifying if the combination C is assigned to buyer B_i,

$y_h(C)$ is a binary variable specifying if the combination C is bought from seller S_h.

$$\sum_{i=1}^{n} \sum_{C \subseteq R} p_i(C) x_i(C) - \sum_{h=1}^{m} \sum_{C \subseteq R} c_h(C) y_h(C) \to \max$$

$$\text{subject to } \sum_{i=1}^{n} \sum_{C \subseteq R} x_i(C) \le \sum_{h=1}^{m} \sum_{C \subseteq R} y_h(C) \ , \ \forall j \in R, \tag{16}$$

$$x_i(C) \in \{0,1\} \ , \ \forall C \subseteq R \ , \ \forall i \ , \ i = 1,2,...,n \ ,$$

$$y_h(C) \in \{0,1\} \ , \ \forall C \subseteq R \ , \ \forall h \ , \ h = 1,2,...,m \ .$$

The objective function expresses the profit of the auctioneer (supply network). The constraints ensures for buyers to purchase a required item and that the item must be offered by sellers. The formulated combinatorial double auction can be transformed to a combinatorial single-sided auction.

Multi-round iterative auctions can be taken as a solution approach. In the fundamental work [1], Bikhchandani and Ostroy demonstrate a strong interrelationship between the iterative auctions and the primal-dual linear programming algorithms. A primal-dual linear programming algorithm can be interpreted as an auction where the dual variables represent item prices. The algorithm maintains a feasible allocation and a price set, and it terminates as the efficient allocation and competitive equilibrium prices are found. For the problem we will formulate the LP relaxation and its dual. Several auction formats based on the primal-dual approach have been proposed in the literature. Though these auctions differ in several aspects, the general scheme can be outlined as follows:

a) Choose minimal initial prices.
b) Announce current prices and collect bids. Bids have to be higher or equal than the prices.
c) Compute the current dual solution by interpreting the prices as dual variables. Try to find a feasible allocation, an integer primal solution that satisfies the stopping rule. If such solution is found, stop and use it as the final allocation. Otherwise update prices and go back to b).

6 Conclusions

The paper proposes and compares coordination mechanisms based on game, contract and auction models. The game theory approach seems to be a basis for the other two approaches. Biform games are promising for combination of non-cooperative and cooperative behaviour of the participants. It may form a global view of the coordination problem. Contracts and auctions are possible to model as specific game theory models. The proposed contract model with stochastic price-dependent demand is a generalized newsvendor model. The proposed combinatorial double auction model is binary linear programming model solved by iterative process. The proposed coordination mechanisms were tested on examples of supply networks.

Combination of these approaches can be a powerful instrument for coordination in supply networks. Coordinating actions in environmental problems is an example of practical using of these approaches. An authority and polluters are looking for coalition projects to reduce pollution (see [17]).

Acknowledgements

This paper was written with the support of the Czech Science Foundation project No. P402/12/G097 DYME - *Dynamic Models in Economics* and by the grant No. IGA F4/57/2017, Faculty of Informatics and Statistics, University of Economics, Prague. The paper is a result of institutional research project no. 7429/2018/08 supported by University of Finance and Administration, Prague.

References

[1] Bikhchandani, S., & Ostroy, J. M. (2002). The package assignment model. *Journal of Economic Theory*, 107 (2), 377–406.
[2] Brandenburger, A., & Stuart, H. (2007). Biform games. *Management Science*, 53(4), 537-549.
[3] Cachon, G. (2003). Supply Chain Coordination with Contracts. In Graves, S., De Kok, T. (eds) *Handbooks in Operations Research and Management Science: Supply Chain Management* (pp. 227-339).Amsterdam: Elsevier.
[4] Cachon, G., & Netessine, S. (2004). Game Theory in Supply Chain Analysis. In: Simchi-Levi, D., Wu, S. D., Shen, M. (Eds.) *Handbook of Quantitative Supply Chain Analysis: Modeling in the e-Business Era* (pp. 13-65). Boston: Kluwer.
[5] Cramton, P., Shoham, Y., & Steinberg, R. (Eds.). (2006). *Combinatorial Auctions*. Cambridge: MIT Press.
[6] Fiala, P. (2005). Information sharing in supply chains. *Omega: The International Journal of Management Science*, 5(33), 419-423.
[7] Fiala, P. (2015): Biform games in supply chains. In: *Mathematical Methods in Economics 2015* (pp. 177–183). Pilsen: University of West Bohemia.
[8] Fiala, P. (2016). Profit allocation games in supply chains. *Central European Journal of Operations Research*, 24(2), 267-281.
[9] Fiala, P. (2016). Supply chain coordination with auctions. *Journal of Business Economics*, 86(1–2), 155–171.
[10] Fiala, P. (2017) Modelling and Solving of Multidimensional Auctions. *Ekonomický časopis*. 65(3), 220–236.
[11] Harrison, T. P., Lee, H. L., & Neale, J. J. (2003). *The Practice of Supply Chain Management: Where Theory and Application Converge*. Boston: Kluwer.
[12] Nagarajan, M., & Sošić, G. (2008). Game-Theoretic Analysis of Cooperation among Supply Chain Agents: Review and Extensions. *European Journal of Operational Research*, 187(3), 719-745.
[13] Shapley, L. S. (1953). A value for n-person games. In: Tucker, A. W., Luce, R. D. (Eds.) *Contributions to the Theory of Games* II (pp. 307-317). Princeton: Princeton University Press.
[14] Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (1999). *Designing and Managing the Supply Chain: Concepts, Strategies and Case studies.* Boston: Irwin/ Mc Graw-Hill.
[15] Simchi-Levi, D., Wu, S. D., & Shen, M. (Eds.). (2004). *Handbook of Quantitative Supply Chain Analysis: Modeling in the e-Business Era.* Boston: Kluwer.
[16] Snyder, L.V., & Shen, Z. J. M. (2011). *Fundamentals of Supply Chain Theory*. Hoboken: Wiley.
[17] Šauer, P., Fiala, P., Dvořák, A., Kolínský, O., Prášek, J., Ferbar, P., Rederer, L. (2015). Improving Quality of Surface Waters with Coalition Projects and Environmental Subsidy Negotiation. *Polish Journal of Environmental Studies*, 24(3), 1299–1307.
[18] Tayur, S., Ganeshan, R., & Magazine, M. (1999). *Quantitative models for supply chain management*. Boston: Kluwer.

Getis' non-parametric approach to spatial data analysis, with application to regional unemployment dynamics

Tomáš Formánek[1], Roman Hušek[2]

Abstract. This contribution focuses on key estimation and stability issues related to spatially augmented econometric models. We describe and evaluate the non-parametric spatial filtering methodology by Getis, which is a specific approach towards quantitative analysis performed using regional (spatial) data. Getis approach is compared to standard parametric spatial models, their differences and similarities are discussed. We use Getis-type econometric framework to estimate and evaluate an empirical model describing regional unemployment dynamics at the NUTS2 level for the following countries: Austria, Belgium, Czechia, Denmark, Germany, Hungary, Luxembourg, the Netherlands, Poland and Slovakia. Alternative neighborhood definitions are used for data filtering, model estimation and stability assessment. In the application part of our contribution, we find prominent evidence for regional influences in the observed data that should be accounted for in quantitative analyses. At the macroeconomic policy-making level, our findings highlight the importance of regional (cross-border) cooperation.

Keywords: spatial data, non-parametric filtering, regional unemployment.

JEL classification: C23, C31, C52, E66
AMS classification: 91B72

1 Introduction

As quantitative macroeconomic analyses are often performed at a regional level, statistical effects of spatial dependency in observed data should be addressed. Spatial autocorrelation violates basic assumptions of independence among observations of variables used for estimation of regression models (see [9] for details). As discussed by [2], [3], [6] and by numerous other authors, uncontrolled spatial autocorrelation leads to biased model estimates and/or biased significance statistics. A common econometric practice is to directly (parametrically) estimate spatial dependencies - either by maximum likelihood methods as in [2] or through a Bayesian approach described in [6]. Unfortunately, parametric approaches rely heavily on prespecified distributional assumptions, see [8].

An alternative approach to the parametric spatial-dependency modeling is outlined and evaluated in this paper: we use a distribution-free nonparametric approach pioneered by Getis [5], which "filters out" spatial dependencies from the data. The theoretical part of our paper is accompanied by an empirical application focused on model stability under varying neighborhood definitions. In our application example, we describe regional unemployment dynamics at the NUTS2 level for the following countries: Austria, Belgium, Czechia, Denmark, Germany, Hungary, Luxembourg, the Netherlands, Poland and Slovakia. Figure 1 is provided for basic illustration of regional unemployment, its spatial distribution and regional dependencies. While this paper outlines and evaluates one possible approach to macroeconomic modeling, other paradigms are certainly relevant as well: see e.g. [1], [3] or [4] for discussion and additional references.

The remainder of this paper is structured as follows: Section two covers the theory of nonparametric spatial filtering and presents references to key literature. Section three provides an empirical application to the topics outlined, along with analysis of model stability. Section four and the list of references conclude our paper.

2 Getis data-filtering in spatial econometric models

When dealing with parametric spatial models, we have to consider the following three different spatial interaction types: spatial interaction effects among observations of the endogenous (dependent) variable, interaction effects among regressors and interactions among error terms. Using a modified notation from [2], we can generalize linear regression models such as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

[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

into a fully spatial specification of a linear regression model (cross-sectional data) as follows:

$$\begin{aligned} \boldsymbol{y} &= \lambda \boldsymbol{W}\boldsymbol{y} + \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{u}, \\ \boldsymbol{u} &= \rho \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}, \end{aligned} \tag{2}$$

where $\boldsymbol{y}$ is a $(N \times 1)$ vector of dependent variable observations, α is the intercept and $\boldsymbol{\iota}$ is a $(N \times 1)$ "unit" vector (all elements are equal to 1), $\boldsymbol{X}$ is a $(N \times k)$ matrix of exogenous regressors; in specification (1), the intercept term $\alpha\boldsymbol{\iota}$ is condensed into $\boldsymbol{X}\boldsymbol{\beta}$. $\boldsymbol{W}$ is a $(N \times N)$ spatial weights matrix as defined below, $\boldsymbol{W}\boldsymbol{y}$ is a spatial lag of the endogenous variable, as defined in [2], [3] or [6], $\boldsymbol{W}\boldsymbol{X}$ is the spatial lag for regressor matrix $\boldsymbol{X}$ and $\boldsymbol{W}\boldsymbol{u}$ describes spatial interactions among disturbance elements $\boldsymbol{u}$. Scalars α, λ and ρ as well as the $(k \times 1)$ vectors $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are the parameters of the model to be estimated. Finally, $\boldsymbol{\varepsilon}$ is the $(N \times 1)$ vector of potentially heteroskedastic error elements. Assumptions and methods for model (2) estimation, statistical inference and interpretation are available e.g. from [6]. We should note that there is only one matrix $\boldsymbol{W}$ specification used in the above model. However, the model (2) may be generalized even further by allowing for different $\boldsymbol{W}_a$ matrices for each of the $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{u}$ elements. This may be appropriate for applications where significantly diverse spatial interactions occur - see e.g. [6]. However, in most practical applications we just assume a common $\boldsymbol{W}$ for the whole model. Admittedly, such simplifying restriction is mostly induced by computational demands and identification issues, rather than fully empirically justifiable.

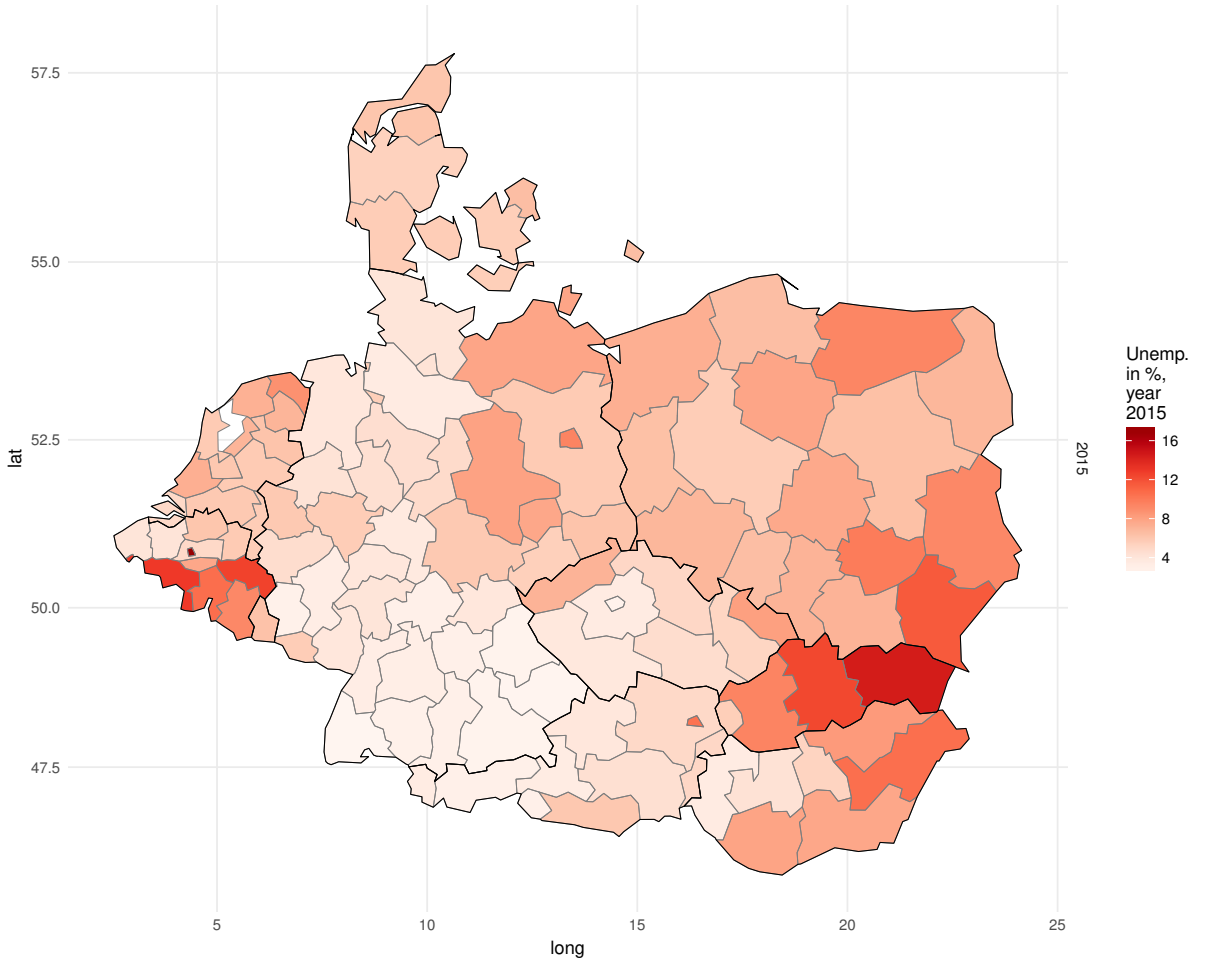


Figure 1 Choropleth of 2015 unemployment rates - NUTS2 level

The spatial weight matrix $\boldsymbol{W}$ from equation (2) is calculated by row-standardizing a spatial matrix $\boldsymbol{S}_N$ $(N \times N)$ that defines neighbors for each spatial unit (region). Neighbor definitions can be produced using dummy variables: the s_{ij} elements of $\boldsymbol{S}_N$ equal 1 if the two spatial units i and j are neighbors and 0 otherwise. Usually, two units are considered neighbors if they are located sufficiently near each other. Diagonal elements of $\boldsymbol{S}_N$ are set to zero by definition (units are not neighbors to themselves). Formally, individual elements of the symmetrical spatial matrix

$\boldsymbol{S}_N$ may be defined as follows:

$$s_{ij}(\tau) = s_{ji}(\tau) = \begin{cases} 0 & \text{if} \quad i = j, \\ 0 & \text{if} \quad d_{ij} > \tau, \\ 1 & \text{if} \quad d_{ij} \leq \tau, \end{cases} \tag{3}$$

where d_{ij} is some adequate measure of distance (aerial km distances are used in this article) between units' representative location points (centroids) and τ is an ad-hoc defined maximum neighbor distance threshold.

Besides computational, robustness and identification issues, the estimated parameters of model (2) cannot be directly interpreted in terms of marginal effects (see [2] for discussion of direct effects and spillovers). Also, the parametric framework discussed above may perform quite poorly under simultaneous presence of different sources of misspecifications. Factors such as unaccounted nonlinear relationship among spatially correlated variables, spatially varying relationships, common factors (spatial and time-related) and spatial heterogeneities can have negative effects on model estimators and their robustness. Fortunately, nonparametric methods may be applied to filter-out spatial dependence from the data and to simplify model estimation and interpretation. Such approach is applicable to positively autocorrelated spatial data (where spatial units with high values of a given variable tend to be surrounded by other high-value observations and vice versa) with a continuous positive support.

Spatial filtering by Getis

Spatial filtering methods may be used to remove global and/or local spatial dependencies among geo-coded variables. Unlike the estimators discussed above, spatial filtering does not rely on distributional assumptions and it is fairly robust to model misspecification. Nonparametric filtering can be used to eliminate spatial autocorrelation from some observed y_i values by "spatial demeaning" through local autocorrelation measures. Spatial filtering method by Getis [5] can be defined as follows: We start with the *Local G*: $G_i(\tau)$ statistic that measures the degree of spatial association. For each observation y_i from a geo-coded sample, we can calculate *Local G* as

$$G_i(\tau) = \frac{\sum_{j=1}^{N} s_{ij}(\tau)\, y_j}{\sum_{j=1}^{N} y_j}, \qquad j \neq i, \tag{4}$$

where $s_{ij}(\tau)$ comes from (3). Observations of variable y are assumed to have a natural origin and positive values. The numerator of (4) is the sum of all y_j observations within distance τ of unit i, but not including y_i. The denominator is the sum of all y_j, not including y_i. Hence, $G_i(\tau)$ is a proportion of the sum of all y_j values that lie within τ distance of i. For example, if we observe high values of y_j within distance τ of unit i, then $G_i(\tau)$ would be relatively high if compared to its expected value under the null hypothesis of full spatial randomness:

$$E\left[G_i(\tau)\right] = \frac{S_i}{N-1}, \tag{5}$$

where S_i is the sum of elements in the i-th row of spatial matrix $\boldsymbol{S}$, i.e. the number of neighbors of i. N is the total number of spatial observations in the sample. For the sake of completeness and under the H_0 of spatial randomness, we can write

$$var\left[G_i(\tau)\right] = \frac{S_i(N-1-S_i)}{(N-1)^2(N-2)} \left(\frac{Y_{i2}}{Y_{i1}^2}\right), \tag{6}$$

where $Y_{i1} = \frac{\sum_j y_j}{N-1}$ and $Y_{i2} = \frac{\sum_j y_j^2}{N-1} - Y_{i1}^2$. Following [7], equation (5) and the square root of (6) can be combined into Ord's G^*, which is a z-score statistic that discerns local spatial clusters (hot-spots/colds-spots) of high/low values of the variable under scrutiny. Finally, a common modification to the $G_i(\tau)$ statistic (not used in this article) consists in dropping the $j \neq i$ restriction from (4). Such *Local G* statistic is denoted $G_i^*(\tau)$ and the values of y_i enter its corresponding numerator and denominator expressions.

The $G_i(\tau)$ statistics from (4) and its expected value $E\left[G_i(\tau)\right]$ from (5) can be used for Getis filtering: the multiplicative transformation of a given geo-coded variable y_i is defined as follows:

$$\ddot{y}_i = \frac{E\left[G_i(\tau)\right]}{G_i(\tau)} \cdot y_i, \tag{7}$$

where $\ddot{y}_i$ is the spatially filtered value of y_i. The transformation outlined in (7) corrects for positive spatial autocorrelation in observed data by counterbalancing the clustering of below-average and above-average observations. Specifically, the filtering factor in (7) shrinks y_i if the majority of observations y_j within the τ distance of unit i are above average. Similarly, y_i is inflated if neighboring observations feature below-average values. As τ increases, the multiplicative transformation element (fraction) in expression (7) approaches unity and its variance decreases.

Should the $G_i^*(\tau)$ statistic be used instead of $G_i(\tau)$, the transformed values $\ddot{y}_i$ would actually equal y_i once τ becomes greater than the distance between i-th region and its most distant counterpart in the sample.

While the filtering approach is computationally simple and intuitive, its underlying assumptions of positive spatial autocorrelation and positive support for y_i can be a limitation. Also, the τ maximum neighbor distance threshold is arbitrary and some level of robustness evaluation is often necessary. Getis filtering as in (7) is univariate. Therefore, if we aim to estimate regression models using spatially filtered data:

$$\ddot{\boldsymbol{y}} = \ddot{\boldsymbol{X}}\boldsymbol{\beta} + \ddot{\boldsymbol{\varepsilon}}\,, \tag{8}$$

then spatial filtering (7) has to be applied individually to each observed variable used in the model ($\ddot{\boldsymbol{y}}$ denotes a vector of spatially filtered dependent variable observations and $\ddot{\boldsymbol{X}}$ is a matrix of spatially filtered regressors).

3 Empirical application

In this section, we outline and estimate a linear regression model describing regional unemployment dynamics. A relatively simple, yet informative and theoretically well-established model (see [3] for discussion) is defined in equation (9), explaining unemployment dynamics in terms of its key determining factors: we use GDP per capita, and two convenient economic structure and competitiveness indicators: the high-tech sector to total labor force ratio and an analogous ratio for the sector of services (with a high proportion of unskilled labor input and predominant non-exportable i.e. non-tradable output). Our model is defined as follows:

$$UN_i = \beta_0 + \beta_1\, GDP_i + \beta_2\, HTC_i + \beta_3\, Services_i + \varepsilon_i\,, \tag{9}$$

where UN_i is the overall unemployment rate observed in the i-th spatial unit (specifically, we use NUTS2 unemployment data from the Eurostat's "lfst_r_lfu3rt" dataset), GDP_i is the region's log-transformed GDP (euro per capita, "nama_10r_2gdp" dataset), HTC_i describes the percentage of employees working in the high-tech industry (NACE r.2 code HTC, "htec_emp_reg2" dataset), $Services_i$ is the proportion of employees within the aggregated services sector (NACE r.2 code H-U, "htec_emp_reg2" dataset), β_j are model parameters to be estimated and ε_i is the random error (see [9] for detailed description of model assumptions). Equation (9) is estimated using 2015 data, based on a total of 111 NUTS2 regions (see figure 1) from Austria, Belgium, Czechia, Denmark, Germany, Hungary, Luxembourg, the Netherlands, Poland and Slovakia. We use data for the year 2015, which provides the most recent and *complete* dataset available for the variables outlined in equation (9). Although observations in some spatial data series featured in model (9) may cover the period from 1999 to 2017, complete datasets for 2016 are not yet available from Eurostat at the NUTS2 level (as of March 2018).

Model type	Coefficients	Estimates	Std. Errors	t-values	Pr(>|t|)
Non-spatial	*(Intercept)*	31.565	3.733	8.457	0.000
linear	*GDP*	-4.046	0.503	-8.041	0.000
model	*HTC*	-0.208	0.139	-1.498	0.137
specification	*Services*	0.240	0.036	6.659	0.000
Data	*(Intercept)*	7.513	6.740	1.115	0.268
spatially	*GDP*	-1.088	0.782	-1.390	0.167
filtered with	*HTC*	-0.354	0.148	-2.399	0.018
$\tau = 165$ km	*Services*	0.161	0.043	3.785	0.000
Data	*(Intercept)*	11.518	7.026	1.639	0.104
spatially	*GDP*	-1.629	0.805	-2.023	0.046
filtered with	*HTC*	-0.311	0.141	-2.210	0.029
$\tau = 217$ km	*Services*	0.180	0.043	4.224	0.000

Table 1 Estimated model, alternative spatial setups used

Before actual model estimation, we begin by evaluating assumptions for the Getis filtering method: All the observed variables as per equation (9) exhibit positive support for our 111-unit sample. Also, all variables are positively spatially autocorrelated at the 5% significance level as tested using the Moran's I statistic (see [3] for testing methodology). Due to space limitations, individual Moran's I test results are omitted here. All relevant statistics and model estimation outputs are available from the authors upon request, along with raw data and R-codes.

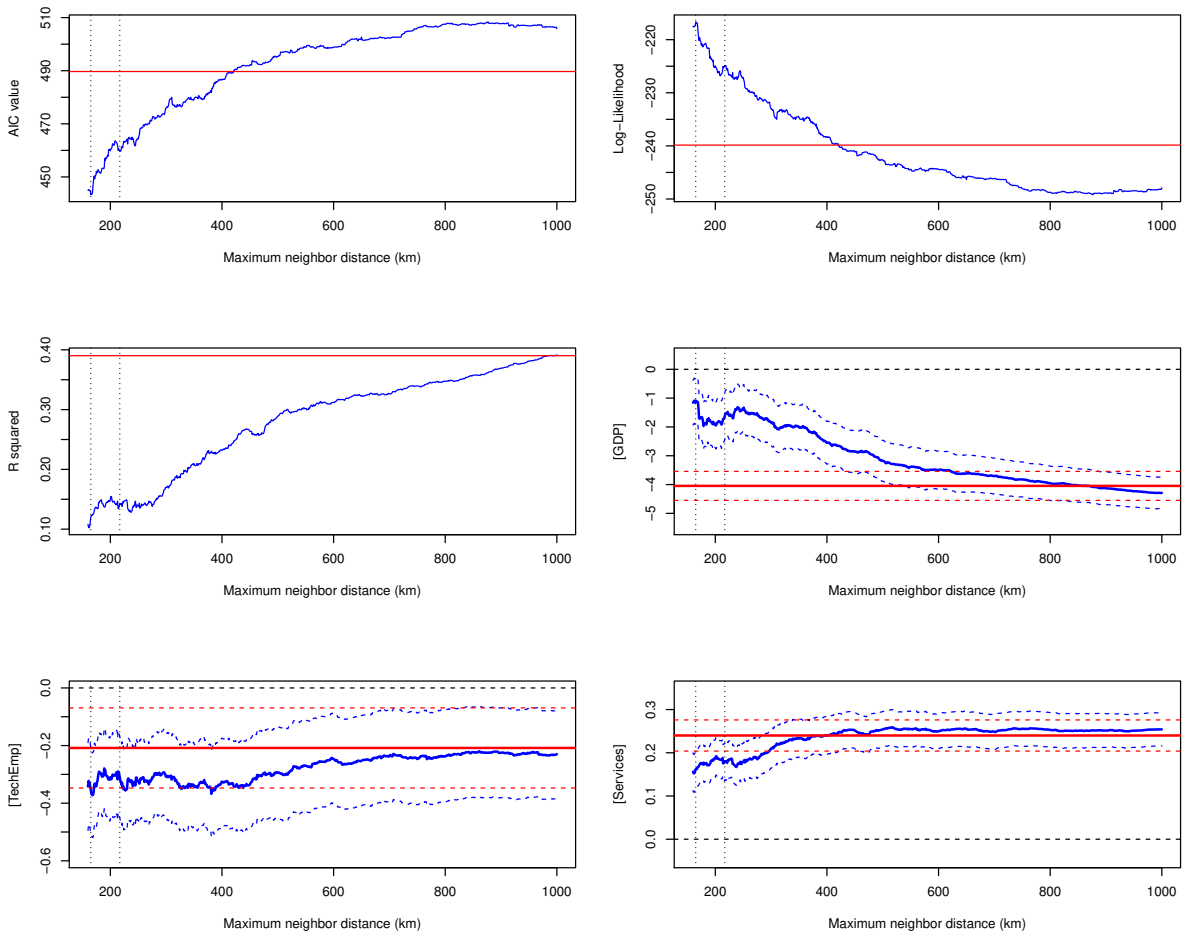


Figure 2 Stability analysis of the model

Table 1 shows key estimation outputs for our model (9). Model estimation output is provided in three versions: first, model (9) is estimated using spatially unfiltered data - output from a linear regression for a non-spatial model specification is shown. Second, spatially filtered data are used - model (9) is estimated using spatially demeaned/filtered data as in equation (8), with maximum neighbor distance threshold τ set to 165 km. This threshold was set empirically, based on the log-likelihood (LL) statistic (LL has indicative properties only, please see following paragraphs for discussion of limits of the LL-based model comparison within the Getis filtering paradigm). The third estimation output also features spatially filtered data. Here, the τ = 217 km threshold relates to an interesting locally optimal spatial setup (see figure 2 and - more importantly - it closely matches our previous findings obtained for a spatial lag model defined in terms of equation (2), as in [3].

Before discussing the estimation results as shown in table 1, we need to address spatial setups and the range of τ distance thresholds used for estimation. In order to evaluate model robustness and statistical properties under different spatial settings, we have estimated our equation (9) using spatially filtered data over a very extensive range of distance thresholds τ: We start from a sparse spatial matrix $\boldsymbol{S}_N$, constructed for $\tau = 160$ km (lower τ values result in "islands" - regions without neighbors - that are incompatible with the logic of Getis filtering). Next, τ thresholds are iteratively increased by a 1-km step up to a maximum neighbor distance threshold of 1.000 km. Although the $\tau = 1.000$ km spatial setup is well beyond reasonably assumed spatial interactions of unemployment dynamics, it provides a nice illustration and comparison between spatial filtering and non-spatial estimation. Overall, a total of 841 spatial specifications of our model (9) in its spatially filtered general form (8) were estimated. Those are summarized in figure 2 as follows: For each τ, we show corresponding Akaike information criteria (AIC), LL, R^2 (see [9] for discussion) and β_j estimates (intercept excluded) along with their $\pm$ one standard error bands. Results from spatially filtered models are shown in blue, while the non-spatial estimation is shown in red for comparison (non-spatial estimate remains constant with respect to changing τ values). Please note that the AIC, LL and R^2 statistics are shown for illustrative purposes only - as spatially filtered (demeaned)

values of the dependent variable differ across alternative τ thresholds (i.e. between estimated models), statistics from different equations should not be directly compared against each other. Finally, estimated values for $\tau = 160$ km and 217 km are highlighted by vertical dashed lines.

Results from table 1 and figure 2 provide several interesting insights into unemployment dynamics and its regional aspects. First of all, the expected effect of GDP and other explanatory variables on unemployment is significantly reduced once regional aspects are considered. This attenuation effect is most prominent for smaller τ values, where spatial filtering is concentrated to a compact neighborhood. This result reflects the fact that unemployment spillovers are often observed among contiguous (spatially adjacent) or otherwise very close neighbors. Overall, most of the statistically significant differences between spatially filtered and unfiltered estimates disappear once τ approaches 500 km. Such results are in line with the theoretically presumed upper bound for unemployment spillover dynamics (spatial interdependence) where work commuting patterns and similar factors play a significant role. Please note that distances between regions are measured using centroids and actual inter-regional commuting distances may be significantly lower than the representative distances used in expression (3).

The estimated coefficients of our empirical model are in line with theoretical expectations: we find evidence supporting an inverse relationship between unemployment and GDP growth. Also, prominent high-tech sector seems to lower the unemployment rate (at the 5% significance level, this effect is statistically significant in spatially augmented models only). Finally, we should stress out the fact that the estimated coefficients from a spatially filtered model such as (8) may be directly interpreted, unlike the $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ coefficients from (2), which do not constitute the usual marginal effects.

4 Conclusions

Spatially augmented models provide an important analysis framework where spatial patterns can be controlled for in macroeconomic analyses. As spatial dependencies are taken into account, we can see that many coefficients in spatially augmented models are attenuated. Along with the detected positive spatial dependence in observed data, such results have two main interpretations: First, there is a prominent influence of regional and potentially cross-border (international) factors in observed macroeconomic data that should be accounted for in quantitative analyses. Second, our results underline the importance of regional (cross-border) cooperation in macroeconomic policies - while this paper focuses on unemployment dynamics, regional interactions and spillovers are present in most macroeconomic variables and processes. Overall, spatial filtering provides a relatively simple and interpretable toolbox for regional (spatial) analyses for all sorts of variables.

Acknowledgements

Geo-data source: GISCO - Eurostat (European Commission), Administrative boundaries: © EuroGeographics. Supported by the grant No. IGA F4/58/2017, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Čížek, O.: Unemployment and Aggregate Demand in Spain, *Ekonomický časopis* **65**(5), (2017), 459–481.

[2] Elhorst, J. P.: *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer-Briefs in Regional Science, Berlin, Springer, 2014.

[3] 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.

[4] Frýd, L.: Granger Causality in Macroeconomics Panel Data. In: Mathematical Methods in Economics. In: *Mathematical Methods in Economics* (A. Kocourek, and M. Vavroušek, eds.), TU Liberec, Liberec, 2016, 225-–230.

[5] Getis, A.: Screening for spatial dependence in regression analysis, *Papers of the Regional Science Association* **69**, (1990), 69–81.

[6] LeSage, J. P., and Pace, R. K.: *Introduction to spatial econometrics*. CRC Press, Taylor & Francis Group, Boca Raton, 2009.

[7] Ord, J. K., and Getis, A.: Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis* **27** (1995), 286—306.

[8] Tiefelsdorf, M., and Griffith, D.: Semiparametric filtering of spatial autocorrelation: the eigenvector approach, *Environment and Planning* **39** (2007), 1193—1221.

[9] Wooldridge, J. M..: *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, 2010.

Valuation of bonds on the international bond markets

Blanka Francová[1]

Abstract. The paper investigates the link between various economic variables and monthly bond return. We define the factors influencing the bond price and its effect according time series before crisis, during crisis and after crisis on the U.S. market. The further analysis show that the exchange risk is important factor affecting bond yield. For this reason, we apply the International Arbitrage Pricing Theory witch decomposition total asset returns to non-currency returns and currency returns. Currency movements affect assets factor loadings and associated risk premiums. Perform tests using corporate bond returns are provided in the period 2001-2017.

Keywords: pricing asset, risk premium, international bond market

JEL Classification: G12, G15, G11
AMS Classification: 90C15

1 Introduction

The market capitalization of international bond markets is much larger than that of international equity markets. However, compared to the large body of the literature on international equity market linkages few empirical studies have been carried out of bond systematic risk. The extent of international bond market linkages merits investigation, as portfolio diversification (Abad, Chuliá, Gómez-Puig, 2009). The benefit of our research is the investigate the impact of macroeconomic and market factors on bond return by examining this question in respect of one large developing country, U.S. As one of the largest financial markets in the world. We apply the universal return decomposition to Solnik's International Arbitrage Pricing Theory for corporate bonds. The posits factor loadings are a function of random prices of assets movements and function of random currency movements. This theory decomposes asset return to asset return for all investors and currency returns for investors with different currency. The Arbitrage Pricing Theory is applied on the stock market and it evaluation stocks better than the Capital Asset Pricing Model. In this paper we seek to contribute to this literature by testing the International Arbitrage Pricing Theory on the bond market proposition that currency movements affect systematic risk factors themselves in addition to residual exchange rate risk. Results tend to confirm that the exchange rate is important factor for valuating bonds. The rest of the article is structured as follows: Section 2 is literature review, Section 3 discusses the data and methodology used to measure factors affecting bond returns. Section 4 display the results, section 5 the conclusion.

2 Literature Review

For pricing asset, it is important systematic risk. We use one-factor or multi-factor model according number of uncertainty as risk (Weissensteiner, 2009). The Arbitrage Pricing Theory is multi-factor model. The multi-factor model is applied by Chao (2016), Huang and Kong (2002), among other. Interest rates and inflation influence bond returns (Huang and Kong, 2002; Elton, 2004; Shih-Wei Chao, 2016). The free-risk rate is important factor for pricing bond (Chao, 2016; Huang, 2002). These factors affect the bond yield for long-term period. We use free-risk rate, interest rate and inflation into the basic model for evaluation bonds. In work Elton (2004), the model is used for the valuation bonds considers the absence of arbitrage and the existence of zero risk. The model is suitable for homogeneous bonds. The Huang and Kong (2002) used market specification and equity market for valuation bonds. Chao (2016) explored various macroeconomic variables and bond returns. Some economic variables are conducive to bond return volatility prediction. The forecasting ability is particularly evident at the short end of the term structure. The systematic risk on the bond market is solved by Abad, Chuliá, Gómez-Puig (2009), Viceira (2012) Ludvigson and Ng (2009). Ludvigson and Ng (2009) study a comprehensive list of macroeconomic variables and find that inflation has important forecasting power for the bond risk premium. They apply the Capital Asset Pricing Model on the bond market. The Capital Asset Pricing Model is attractive and simple domestic model that perform covariance of asset returns with state variables. The assumption of the Capital Asset Pricing Model is identified market portfolio. The investors diversify their portfolio across international markets. The exchange rate risks affect returns of international investments (Sirr, 2011, Liu, 2012, Panda, 2013). The International Capital Asset Pricing Model was applied by Solnik (1974), Sercu (1980), exchange rate risk has traditionally been specified as separate factor affecting the prices of capital assets. The International Capital Asset Pricing Model is applied on international market. The market portfolio is efficient does not exist in the international framework. The

[1] Mendel University in Brno, department of Finance, Zemědělská 1, 613 00 Brno, blanka.francova@mendelu.cz

Arbitrage Pricing Theory provides a fruitful alternative to these utilitybased models. In the Arbitrage Pricing Theory, risk factors should explain variation of returns in large (Brennan, 1998). The International Arbitrage Pricing Theory is formulated by Solnik (1983). The form of the International Arbitrage Pricing Theory model is unchanged by investors with different home currencies. Main assumptions of the International Arbitrage Pricing Theory are purchasing power parity, efficient markets and possibility of arbitrage. Solnik (1983) applied the International Arbitrage Pricing Theory as first. The International Asset Pricing Theory decomposes asset returns into portions due to currency returns and non-currency returns. Armstrong (2011) defined changes in underlying asset values (non-currency) and random currency movements (i.e., exchange rate risk). Non-currency returns are earned by all investors regardless of their home currency. The exchange risk is associated with random changes in currency values and change the asset returns of investors with different home currencies. Currency movements affect risk factors and addition residual exchange risk. We apply this universal return decomposition to Solnick's (1983) International Arbitrage Pricing Theory. In this paper, we seek proposition that currency movements affect systematic risk factors themselves in addition to residual exchange rate risk.

3 Methodology and Data

The U.S. bond market is the highest bond market on the world bond market. The U.S bond market is 43 % of the world bond market. The corporate bonds are 85 % of the U.S. bond market. The corporate bonds are the most bond market. Our panel, the U.S. companies issuing corporate bonds, include rich dataset of 71 195 corporate bonds on the U.S. market. Our monthly panel data set covers the period from January 2001 to November 2017. The bond return is yield to maturity. We consider several bond specific variables: issue size and currency of issue. Taking all the restrictions of the different data sources into account we end up with the dataset that contains over 850,000 observations for USA available for the estimations. A detailed description of the data and our sources is given in the Appendix. The main factors affecting bond prices are free-risk rate (Chao, 2016; Abad, Chuliá a Gómez-Puig, 2009), interest rate (Viceira, 2012; Huang, 2002; Brennan, 2001, Boubaker, 2017) and inflation (Brennan, 2001; Brier, 2009). Using the panel dataset, we focus on the impact of factors on the corporate bonds i in time t, where the cross-sectional dimension and the number of time periods. The basic evaluation model is:

$$R_{it} = \alpha_{it} + \beta_1 rf_{\mathrm{t}} + \beta_2 ir_{\mathrm{t}} + \beta_3 if_{\mathrm{t}} + \varepsilon_{it}, \tag{1}$$

where R_{it} is the bond return, α_{it} is a constant, rf_{t} is free-risk rate, ir_{t} is interest rate, if_{t} is inflation, $\varepsilon_{\mathrm{it}}$ is error. Method OLS is applied with fixed effects.

We use the model of the International Arbitrage Pricing Theory. The baseline estimation is specified as (Armstrong, 2011):

$$R_{it} = \alpha_{it} + ER_t + \sum_{b=1}^{3} \beta_b bv_{tb} + \sum_{m=1}^{M} \gamma_m mv_{tm} + \varepsilon_{it,} \tag{2}$$

where ER_t is exchange rate. The variable bv_{tb} denoted the variables of basic model (1), represented free-risk rate, interest rate and inflation. The next variable mv_{tm} denoted further variables, includes exchange rate, risk in the country, bond index and stock index. Possibly heteroscedastic residual, ε_{it}. This model is applied for method OLS with fixed effect estimator.

The exchange rate risk is very significant on the international market for investors with different national currency (Sirr, 2011, Liu, 2012, Panda, 2013). According Armstrong (2011), the International Arbitrage Pricing Theory posits that factor loadings are affected by exchange rate risk. The potentially complex effects of exchange rates on factor loadings via the linear function and the International Arbitrage Pricing Theory model is:

$$R_{it} = \alpha_{it} + ER_t + \sum_{b=1}^{3} \beta_b bv_{tb} + \sum_{s=1}^{2} (\beta_{0s} + \beta_{1s} ER_{0t})\, sv_{ts} + rv_t + \varepsilon_{it,} \tag{3}$$

where sv_{ts} are bond index and stock index, β_{0s} coefficients correspond to the universal component of factor, β_{1s} coefficients capture the exchange risk component of factor arising from issue currency and rv_{tf} is risk. This model is applied for method OLS with fixed effect estimator.

4 Results

This section presents the results for the influence factors on the corporate bond returns. We apply regress analyses for monthly period from 2001 to 2017. We present the results for OLS methods with fixed effects.

We start our analysis with the examination of determinants affecting bond prices. Table 1 presents the results of the first four models that vary according to the inclusions of basic model, model with further factors, model with exchange risk component of factor loadings and model with all variables. The first model (1) includes only basic variables (free-risk rate, interest rate and inflation). Our results show that the coefficient of free-risk rate is positive. The risk-free rate is basic of asset return. The asset returns are decomposed to free-risk returns and returns of risk premiums. The interest rate tends to slow down bond price possibly through reduction of bond yields. The internal value of bonds is decreased by increasing interest rate. The price of bonds depends on its internal value. The inflation is important forecasting power for bond risk premium. If the inflation raises that the nominal bond yield raises too because the investors want to compose a lower real return of bonds. The second model (2) includes further factors (exchange rate, risk, bond index and stock index). These variables are significant. The factors represent further risk premium. If the nominal effective exchange rate increases the national currency depreciates. Depreciates of issuer currency increases bond yield. The impact between stock index and bond yield is negative. If the stock prices decrease than the uncertainty is growing on the capital market. The investors invest into the bond as safely investments and bond prices increase. The third model (3) includes exchange risk factor. The exchange risk component of factor bond index and stock index arising from issue currency. The coefficient of exchange risk is significant as single factor and factor together with market index too. The fourth model (4) includes all variables. Models do not respond by changing the most coefficient estimates after adding additional variables. The models are stable and correctly capture the causality between the variables. These results imply that volatilities of these macroeconomic and market variables could also forecast bond return. The important factor is exchange risk.

	(1)	(2)	(3)	(4)
free-risk rate	0.057***	0.082***	0.168***	0.174***
	(0.017)	(0.017)	(0.017)	(0.017)
interest rate	-2.078***	-2.050***	-1.936***	-1.899***
	(0.030)	(0.030)	(0.031)	(0.031)
inflation	0.142***	0.126***	0.103***	0.100***
	(0.010)	(0.011)	(0.011)	(0.011)
exchange rate		-0.045***		-0.043***
		(0.002)		(0.002)
risk		0.031***		0.136***
		(0.011)		(0.011)
bond index		-0.622***		-0.053
		(0.164)		(0.175)
stock index		-0.144***		-0.058***
		(0.008)		(0.008)
bond index and exchange risk			-0.016***	-0.014***
			(0.001)	(0.001)
stock index and exchange risk			-0.001***	-0.001***
			(0.000)	(0.000)
constant	5.795***	5.821***	5.778***	5.797***
	(0.004)	(0.004)	(0.004)	(0.004)
Observations	826,221	824,560	824,560	824,560
R-squared	0.006	0.007	0.008	0.008
Number of bonds	18,139	18,130	18,130	18,130

Table 1 The main results

The results of Hausman test are in Table 2. The Hausman test shows that we have to use the method OLS with fixed effects. It means that the corporate bonds are not substitutes. The issuer of corporate bonds are financial institutions and non-financial institution. The highest part of the corporate bond market are corporate bonds of financial institutions about 40 % of the U.S. bond market. The non-financial institutions are about 16 % of the U.S. bond market. The last part is government bonds. Our dataset includes financial institutions and non-financial institutions. The corporate bonds rep-resent 56 % of the U.S. bond market.

	fixed effects	random effects	difference
free-risk rate	0.174	0.164	0.010
interest rate	-1.899	-1.916	0.017
inflation	0.100	0.120	-0.020
exchange rate	-0.043	-0.048	0.005
risk	0.136	0.134	0.002
bond index	-0.053	-0.040	-0.013
stock index	-0.058	-0.059	0.001
bond index and exchange risk	-0.014	-0.014	0.000
stock index and exchange risk	-0.001	-0.001	0.000

Test: Ho: difference in coefficients not systematic

chi2(9) = 2044.72	
Prob>chi2 = 0.000	

Table 2 The results of Hausman test

Table 3 shows results by issue size for the corporate bonds. The interest rates have the most impact on bond prices in case small bond issues. The impact of interest rates is decreasing with increasing volume of bond issues. The coefficient of exchange rate is not significant for small issue sizes. The foreign investors buy investment of big companies with high bond issues. The investors with different national currency must follow the development of the exchange rate and currency risk. Therefore, the exchange risk is significant for bonds in higher issues. The bond index and exchange risk are significant for high bond issues.

	(1)	(2)	(3)	(4)
	to 1 million	from 1 million to 10 millions	from 10 millions to 3.5 milliards	from 3.5 milliards
free-risk rate	-0.136***	-0.071	0.257***	0.287***
	(0.036)	(0.053)	(0.015)	(0.040)
interest rate	-3.941***	-2.816***	-1.682***	-1.102***
	(0.510)	(0.215)	(0.137)	(0.232)
inflation	0.066**	0.151***	0.089***	0.120**
	(0.028)	(0.038)	(0.014)	(0.050)
exchange rate	-0.001	-0.034***	-0.053***	-0.071***
	(0.005)	(0.007)	(0.003)	(0.010)
risk	0.027**	0.058***	0.170***	0.151***
	(0.013)	(0.022)	(0.019)	(0.029)
bond index	0.278	1.780***	-0.302	-1.870***
	(0.210)	(0.363)	(0.243)	(0.530)
stock index	0.066***	0.090***	-0.110***	-0.079***
	(0.012)	(0.017)	(0.013)	(0.024)
bond index and exchange risk	0.003	-0.003	-0.017***	-0.011***
	(0.003)	(0.003)	(0.001)	(0.004)
stock index and exchange risk	-0.000***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
constant	6.996***	6.046***	5.537***	5.296***
	(0.006)	(0.008)	(0.004)	(0.018)
Observations	107,169	124,844	571,573	20,974
R-squared	0.017	0.014	0.008	0.018
Number of bonds	3,843	3,509	10,229	549

Table 3 The results by issue size

Because the significance of forecasting variables might change over time, the predictive regressions are also estimated for three shorter horizons. The first one covers from 2001 to 2006, which includes time before crisis. The subsequent 2007-2009 represents period of crisis and 2010-2017 is typical for time of low inflation and low interest rate. The results in times are presented in Table 4. The free-risk rate is not significant in the time after crisis. The coefficient of interest rate is positive in period with low inflation and low interest rate. The coefficient of inflation is negative in before crisis and after crisis periods. The inflation declines real return of bonds. If inflation increases, the real bond return declines. In the crisis, the inflation is positive. The investors expect higher inflation and they want higher bond returns and the bond prices must increase. The coefficient of exchange rate is significant and positive for before crisis and pre-crisis period. In the time with low interest rate and low bond yield, the investors buy bonds with a different target that obtaining a bond yield, for example for currency yield. The coefficient of stock index is positive in pre-crisis and after-crisis period. The capital markets were distrust at the time of the crisis and investors were cautious. The stock index reflect situation on the capital market. If it was successful in stock markets, it was also in confidence in bond instruments. The investors often flow from risky assets to securities with little default risks during financial crisis.

	(16)	(17)	(18)
	from 2001 to 2007	from 2008 to 2009	from 2010 to 2017
free-risk rate	0.320***	0.326***	-0.015
	(0.016)	(0.064)	(0.018)
interest rate	-0.770***	-1.731***	1.013***
	(0.021)	(0.075)	(0.207)
inflation	-0.039***	0.365***	-0.058***
	(0.012)	(0.032)	(0.010)
exchange rate	0.011***	0.023**	-0.031***
	(0.003)	(0.010)	(0.003)
risk	-0.091***	0.249***	0.083***
	(0.014)	(0.045)	(0.010)

bond index	-3.922***	2.762***	-0.434**
	(0.175)	(0.541)	(0.183)
stock index	-0.317***	0.073**	0.016**
	(0.011)	(0.033)	(0.007)
bond index and exchange risk	-0.005***	-0.000	-0.007***
	(0.001)	(0.002)	(0.001)
stock index and exchange risk	0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
constant	6.608***	6.784***	5.451***
	(0.004)	(0.021)	(0.004)
Observations	134,978	105,885	583,697
R-squared	0.029	0.012	0.001
Number of bonds	3,063	3,668	17,497

Table 4 The results for times

5 Conclusions

We applied the International Arbitrage Pricing Model for monthly returns of corporate bonds on the U.S. market for the period 2001-2017. We used bond returns decompose into asset returns and currency returns. Asset returns are earned by all investors regardless of their home currency. Currency returns are earned by investors with different home currencies. The bond returns are yield to maturity. We contribute to the literature on bond return forecast ability by showing that macroeconomic fundamentals have important predictive power for returns on U.S. corporate bonds. The results of the paper are the following. First, the results show that the variables in the basic model are important (free-risk rate, interest rate and inflation). The models are stable and correctly capture the causality between the variables. The corporate bonds are not substitutions and the bonds of different companies behave differently and must be considered separately. Second, we analyzed the exchange risk as risk premium for pricing bond. The exchange risk is significant for most observations. The exchange risk is not significant for small bond issues. Empirical tests confirm that currency movements affect factor loadings and risk premiums. We interpret the results to mean that the total exchange rate risk of bonds is composed of the sum of exchange risk contained in the market factor loading plus residual market factor. Data support a predictive relationship between bond yields and several economic variables in a certain historical period.

Acknowledgements

This research was funded by the Czech Science Foundation, grant No. PEF_DP_2018015 of the topic "Valuation of bonds on international markets ".

References

ARMSTRONG, Will J., Johan KNIF, James W. KOLARI and Seppo PYNNÖNEN. Ex-change risk and universal returns: A test of international arbitrage pricing theory. Pacific-Basin Finance Journal. 2011, 20(1), 24-40. DOI: 10.1016/j.pacfin.2011.08.003. ISSN 0927538x.

ABAD, P., CHULIÁ, H., GÓMEZ-PUIG, M. 2009. EMU and European government bond market integration. Journal of Banking & Finance, 34(12), 2851-2860.

BRENNAN, M.J., CHORDIA, T., SUBRAHMANYAM, A. Alternative factor specifica-tions, security characteristics, and the cross-section of expected stock returns. Journal of Financial Economics 49, 345-373, 1998

ELTON, EDWIN J., MARTIN J. GRUBER, DEEPAK AGRAWAL A CHRISTOPHER MANN. Factors affecting the valuation of corporate bonds. Journal of Banking & Fi-nance. 2004, 11(28), 20.

HUANG, JING-ZHI, AND WEIPENG KONG. "Explaining credit spread changes: Some new evidence from option-adjusted spreads of bond indices." Stern School of Business Working Paper No. FIN-03-013 (2003).

CHAO, Shih-Wei. Do economic variables improve bond return volatility forecasts? Interna-tional Review of Economics and Finance. 2016, 46, 10-26. DOI: 10.1016/j.iref.2016.08.001. ISSN 10590560.

CHRISTENSEN, JENS H. E. A GLENN D. RUDEBUSCH. The Response of Interest Rates to US and UK Quantitative Easing. The Economic Journal. 2012, 122(564), F385-F414

LUDVIGSON, Sydney C. a Serena NG. Macro Factors in Bond Risk Premia. Review of Fi-nancial Studies. 2009, 22(12), 5027-5067. DOI: 10.1093/rfs/hhp081. ISSN 0893-9454.

SERCU, P. Ageneralization of the international asset pricing model. Revue de I'Association Francaise de Finance 1, 91-135. 1980

SIRR, GORDON, JOHN GARVEY A LIAM GALLAGHER. Emerging markets and port-fo-lio foreign exchange risk: An empirical investigation using a value-at-risk decom-po-sition technique. Journal of International Money and Finance. 2011, 30(8), 1749-1772. DOI: 10.1016/j.jimonfin.2011.08.002. ISSN 02615606.

SOLNIK, BRUNO. International Arbitrage Pricing Theory. Journal of Finance. 1983, 38(2), 449-457. ISSN 00221082.
VICEIRA, L. M. 2012. Bond risk, bond return volatility, and the term structure of interest rates. International Journal of Forecasting, 28(1), 97-117.
REZENDE, R. B. 2017. The interest rate effects of government bond purchases away from the lower bound. Journal of International Money and Finance. 165-186.
WEISSENSTEINER, ALEX. 2009. Using the Black-Derman-Toy interest rate model for port-folio optimization. European Journal of Operational Research. 2010(202), 6.

Appendix

Name of variable	Description	Source
yield to maturity	Total yield of corporate bond to maturity of bond	Morningstar
interest rate	Money market rate	International Monetary Fund
inflation	Costumer price index	International Monetary Fund
exchange rate	Nominal effective exchange rate	World bank
free-risk rate	Return of treasury bill	Morningstar
risk	Economic Policy Uncertainty Index	Economic Policy Uncertainty
bond index	U.S. Bond Index Premium	Finance yahoo
stock index	Standard & Poor 500 stocks	Finance yahoo

Table 5 Definition of analyzed variables

Variable	Obs	Mean	Std.Dev.	Min	Max
yield to maturity	864,926	5.846	4.371	-19.587	1295.544
free-risk rate	1,027,084	-0.010	0.232	-1.310	1.200
interest rate	14,381,390	-0.025	0.172	-0.960	0.280
inflation	14,381,390	0.165	0.379	-1.902	1.101
exchange rate	14,356,259	-0.011	1.798	-47.440	20.630
risk	14,381,390	-0.002	0.384	-1.355	1.774
bond index	1,025,327	0.000	0.024	-0.099	0.079
stock index	1,025,327	0.094	0.560	-1.973	1.594
bond index and exchange risk	1,025,055	0.117	4.688	-51.901	44.454
stock index and exchange risk	1,025,055	-21.774	129.978	-2005.652	1680.571

Table 6 Descriptive statistics

	yield to maturity	free-risk rate	interest rate	inflation	exchange rate	risk	bond index	stock index	bond index and exchange risk
free-risk rate	-0.006	1							
interest rate	-0.067	-0.017	1						
inflation	0.023	-0.097	0.180	1					
exchange rate	-0.054	-0.128	0.021	-0.001	1				
risk	0.004	-0.089	-0.009	-0.013	-0.002	1			
bond index	-0.005	-0.037	0.079	-0.016	0.027	0.110	1		
stock index	-0.029	0.185	0.096	-0.097	-0.260	-0.079	-0.191	1	
bond index and exchange risk	-0.024	0.077	0.049	-0.069	0.213	-0.047	0.361	-0.114	1
stock index and exchange risk	-0.044	0.155	0.150	-0.054	-0.156	0.255	-0.091	0.390	-0.242

Table 7 Correlation matrix

Stochastic modelling of provisions in company providing loans – case study

Ludvík Friebel[1], Jana Friebelová[2],

Abstract. Calculation of provisions is a very important task for each company providing credit loans. Two methods of calculation are used for this purpose. The method based on expected credit loss is commonly used. Expected loss is (in a simplified way) computed as a product of outstanding, probability of default and loss given default. In our contribution we will use the model based on Markov chains for the same purpose. Provisions to the realised contracts are depending on many variables, for example type of loans, type of client, duration, number of instalments from start, number of instalments to end. Overall economic situation has got considerable influence too. In this article, one part of provisioning model – probability of termination - and two approaches to it were analysed. In conclusion, a comparison between commonly used model and model based on simulation is done.

Keywords: provisions, consumer loan, expected credit loss, default, lost given default, probability of termination.

JEL Classification: C44

AMS Classification: 90C15

Introduction

The financial crisis of 2008-2009 revealed that improper estimation of credit risk leads to dramatic impact on the economy. Various credit risk models developed to estimate credit risk is presented by Bluhm et al. [1], Duffie [2], Giesecke [3], Lando [5], McNeil [6], In work of Valle [11], there are identified three methodologies for credit risk modelling: structural form models (SFM), reduced form models (RFM) and factor models (FM). SFM are based on the Black and Scholes theory for option pricing and on the Merton model. Linda [6] has identified two different methodologies for modelling credit risk: an options-theoretic structural approach developed by Merton [7] and a reduced form approach utilising intensity-based models to estimate stochastic hazard rates. However, they both agree that the structural approach models the economic process of default, but reduced form models decompose risky debt prices to estimate the random intensity process underlying default. Therefore, RFM mainly focuses on the accuracy of the probability of default (PD), such that it is more important than an intuitive economical interpretation. More realistic credit spreads can be obtained from reduced form models (RFM) or intensity-based models (Linda [5]).

A quantification method for the collection effect on consumer term loans is dealt with by Ping et al. [9]. Modelling loss given default using quantile regression is introduced by Krüger and Rösch [4]. Comparison of current credit risk models is given by Spuchľakova and Cug [10].

Simplified Methodology of provisioning

Amount of provisions for particular portfolio is given due parameter ‚provisioning rate'. This parameter gets value 0-100% and determines rate of covering of outstanding by provisions. Provisioning rate (PR) is function of following variables:

- PD = probability of default,
- LGD_1 = probability of termination defaulted contract,
- LGD_2 = rate of loss on terminated contract.

Each of these variables can be defined at contract level. Their estimates will always be made on a set of contracts that are homogeneous. PR is the same for all contracts in one given pool.

[1] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, České Budějovice, friebel@ef.jcu.cz.
[2] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, České Budějovice, friebelova@ef.jcu.cz.

Aggregation criteria are:

- type of portfolio (common, in insolvency, high risk);
- type of product (Cars, Credit cards, Consumer goods etc.);
- days past due (DPD) or information about termination (DPD = 0, 1-30, 31-60, 61-90, 91+ or status of termination);
- number of quarters from date of termination.

PR is calculated at the contracts level according to formula:

$$PR = PD \text{ x } LGD_1 \text{ x } LGD_2, \tag{1}$$

where PD, LGD_1 and LGD_2 are estimated for given pool.

With regards to data availability, formula (1) is used in the form

$$PR = PT \text{ x } LGD_2, \tag{2}$$

where probability of termination $PT = PD \text{ x } LGD_1$ in case that PD is not monitored.

The move from active contracts to termination within a given horizon, a variable that interprets $PD \text{ x } LGD_1$. We will write a $PT(k, h)$, where k is the interval of days past due, and h is the horizon under which contracts are terminated. The LGD_2 parameter then corresponds to LGD_2 for the "newly" canceled contracts on the given portfolio, i.e. the number of quarters from time of termination = 0. So, we will write $LGD_2(0)$.

According to IFRS9 methodology valid since 1. 1. 2018 non-default portfolio (i.e. healthy: contracts with 0-30 DPD and watch contracts: contract with 31-90 DPD) is provisioned, resp. expected credit loss (ECL) is calculated. Formula for calculation of ECL is more sophisticated. One part of this formula is still PT, is calculated till 12 months or to maturity of contract and is expressed by PT curves.

Determination of PT based on historical data - portfolio approach

PT curves are calculated on historical data like probability of termination in given quarter for each product line and stage. Portfolio approach means that in each input quarter there are contracts in different phase of life and just rest of life to maturity is monitored.

Q_0	Q_1	Q_2	…	Q_{24}
03/2012	x	x	x	x
06/2012	x	x	x	
…	x	x		
12/2017	x			

Table 1 Sample table

Repaid and contentious contracts are removed for next quarter. For example, in quarter Q_0 entry 100 contracts and during one quarter 5 contracts were contentious and 10 contracts repaid. Probability of termination during first quarter is 5/100. Next quarter entry 85 contracts and 5 were repaid and 5 were contentious – then PT=5/85 and so on. Average probability for each quarter is available.

Non-default active contracts with at least one instalment in future from one product line with total number of instalments=24 were chosen from portfolio (input period Q_0 from 03/2012 to 12/2017). For purpose of this paper, probability of termination till maturity on two groups (healthy and watch) and consequently average PT for Q_1 to Q_{24} were calculated.

Cumulative PT (from average quarterly PT) calculated as

$$PT(Q_1)+(1-PT(Q_1))\cdot PT(Q_2)+(1-PT(Q_1))\cdot(1-PT(Q_2))\cdot PT(Q_3)+(1-PT(Q_1))\cdot(1-PT(Q_2))\cdot(1-PT(Q_3))\cdot PT(Q_4) \ldots + \ldots \tag{3}$$

is used (together with other parameters) for estimation of ECL for given group.

Determination of PT by simulation – vintage approach

Unlike portfolio approach, vintage approach considers both duration of contract from date of signature up to the monitored period and duration since monitored period till maturity too.

The data file consisted of 6595 records of contracts represent consumer loans with 24 instalments concluded in year 2015. Each record included DPD or status of contract in particular month from contract start date. The data had to be processed to gain necessary simulation inputs, see below.

Contract can be classified into seven separate groups according to DPD or statuses representing Default, Termination or Repayment before maturity see Table 2. Groups according to DPD and Default represent active contracts. Contracts are terminated due two reasons: DPD > 150 days or extra status (insolvency, fraud, death).

DPD/status	0	1-31	31-60	61-90	Default	Terminated	Repaid
j	0	1	2	3	4	5	6

Table 2 Status of contract

Simulation model

Simulation inputs:

- Number of months from contract start date i=0, …, n;
- k is situation on contract in month i;
- l is situation on contract in month $i+1$;
- p_{ikl} is probability that contract in month i transit from situation k to situation l, in our simulation model we use probability mass function (pmf) for each initial situation for each month. We obtained this pmf from data analysis. Particular probabilities for month i can be written into transition matrix $\mathbf{P}_i$.
- T number of simulated contracts for each possible initial situation and each month of contract life.

Simulation outputs:

- $c_{ijklsim}$ number of contracts in month i in situation k which were in month j in situation l;
- $r_{ik5sim}= C_{ik5sim} /T$ total amount of terminated contracts till maturity which were in month i in situation k;
- $r_{ik5sim}= C_{ik5sim} /T$ is total ratio of terminated contracts for month i and situation k;
- $\boldsymbol{r}_{i5sim}$ is vector with elements r_{ik5sim};
- s_{ik} share of contracts in situation k in month i;
- $\boldsymbol{s}_i$ vector with elements s_{ik};
- $\boldsymbol{R}_5$ vector of share terminated contract, where $R_{i5}=\boldsymbol{r}_{i5sim}\,\boldsymbol{s}_i =$ ratio of terminated contracts among i-month old contracts.

For simulation, SW @RISK 7.5 is used from package Palisade Decision Tools. 50000 contracts for all possible situations within period of 24 months was simulated. At the beginning of the contracts, we simulated 50000 contracts only. For the first month after the beginning there are two possible situations: instalment is not delayed or is delayed 1-31 days, $k\in\{0, 1\}$. Therefore, we performed two simulations for the two mentioned situations. In such way we continued up to 24 months. Transition matrix for i=6 is depicted in table 3.

	0	1	2	3	4	5	6
0	0.963	0.025	0.000	0.000	0.000	0.000	0.012
1	0.542	0.251	0.187	0.000	0.000	0.010	0.010
2	0.258	0.194	0.065	0.484	0.000	0.000	0.000
3	0.071	0.000	0.071	0.143	0.607	0.071	0.036
4	0.077	0.000	0.000	0.000	0.308	0.615	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3 Transition matrix $\mathbf{P}_6$

This procedure gave for each possible initial situation of contracts of different vintage a random vector including number of terminated contracts for each remaining month of life of contract - $\boldsymbol{r}_{i5sim.}$ With this vector we can predict the probability of termination for particular contract up to its maturity. In Table, there is shown the termination probability for four-month old contracts in five possible situations. Simulated results give clear characteristic for a longer time unpaid contracts $k \in \{2, 3, 4\}$ where probability of termination is very high.

	Following month (*j*)							
k	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
0	0.000%	0.040%	0.000%	0.040%	0.226%	0.226%	0.122%	0.290%
1	0.000%	0.286%	0.750%	2.442%	3.152%	1.804%	0.548%	0.594%
2	2.040%	3.680%	18.286%	7.718%	6.212%	2.632%	0.588%	0.510%
3	12.000%	36.842%	10.374%	2.778%	1.842%	0.722%	0.210%	0.216%
4	74.194%	7.530%	2.738%	0.896%	0.838%	0.498%	0.118%	0.118%

Table 4 Termination probability for after 4th month, l=5, j=5, …, 12

Results and conclusion

In order to compare both approaches (portfolio app. and vintage app. with simulation) average probability of termination up to maturity of contracts were calculated.

In portfolio approach, all healthy and watch contracts regardless of duration from start date with at least one instalment in future are included into each from 24 input quarter (03/2012-12/2017). Average *PT* for each further quarter till maturity were computed and from these average *PT* cumulative PT were calculated. *PT* of healthy contracts till maturity = 2.9% and *PT* of watch contracts = 25.9%. Rate of healthy contracts = 99.26%, rate of watch contracts = 0.74%. *AVG PT* weighted by rate =3.09%.

Simulation model was based on contracts from one product line with 24 instalments started in year 2015. Repayment of these contracts was monitored and from obtained data the simulation model was proposed. Important difference between portfolio approach and simulation model is in the number of accounts which were taken into account. In each simulation representing different period from the beginning of contract, there was the same number of iterations used – same number of simulated contracts. This procedure assumes always the same number of contracts of different ages. This assumption cannot be fulfilled in real reference sample, where number of contracts decreases with the growing number of months from the start because of termination and repayment before maturity. For reason of comparability of the results obtained with different approach there is needed to use weighted average for computation of *AVG PT* in case of simulation. Elements of vector of weight $\boldsymbol{d}$ is computed following way:

$$d_i = \frac{\sum_{k=0}^{6} s_{ik} - s_{i5} - s_{i6}}{\sum_{i=1}^{n} \sum_{k=0}^{6} s_{ik} - s_{i5} - s_{i6}}, \tag{4}$$

where s_{ik} is number of contracts in portfolio in month i.

There after $AVG\ PT$ obtained with simulation model was computed

$$AVG\ PT = \boldsymbol{d} \cdot \boldsymbol{R}_5, \tag{5}$$

$AVG\ PT$ for simulation model was 3.04 %, which is a bit less than the corresponding value obtained with portfolio approach.

Portfolio approach used in this article is more simply compared to vintage simulation model, but results are very similar. The advantage of simulation is taking into account the number of months from the beginning. This feature can be used to determine provisions more precisely for each status of contract in each period of life.

References

1. Bluhm, C., Overbeck, L., Wagner, Ch. (2010). *An Introduction to Credit Risk Modeling, Second Edition*. Boca Raton: CRC Press.
2. Duffie, D. Singleton, K. J. (2003). *Credit Risk: Pricing, Measurement, and Management*. Princeton: Princeton University Press.
3. Giesecke, K. (2004). *Credit Risk: Models and Management: 2nd Edition*. Risk Books, 2, 487-526.
4. Krüger, S., Rösch, D. (2017). Downturn LGD modeling using quantile regression. *Journal of Banking and Finance 79*. 42-56.
5. Lando, D. (2004). *Credit Risk Modeling: Theory and Applications*. Princeton: Princeton University Press.
6. Linda, A. P. (2004). *Credit Risk Modeling of Middle Markets*. New York: Zicklin School of Business, Baruch College, CUNY.
7. McNeil, A. J., Frey, R., Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton: Princeton University Press.
8. Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, *29(2)*, 449-470.
9. Ping H., Zhongsheng, H., Zhixin, L. (2015). A quantification method for the collection effect on consumer term loans is depicted. *Journal of Banking & Finance 57*, 17-26.
10. Spuchľakova, E., Cug, J. (2015) Credit Risk and LGD Modelling. *Procedia Economics and Finance 23*, 439-444.
11. Valle, A. C. (2013). *Credit Risk Modelling in a Semi-Markov Process Environment*. Manchester: Manchester University; School of Mathematics.

The Value Premium of Skewness in the Existence of Cross-section Dependence

Lukáš Frýd [1]

Abstract. The main contribution of this paper is to the broader understanding of risk premium in the financial stocks of $S\&P\ 500$. We utilize the model based on the arbitrage pricing theory (APT) with the two source of risk, variance, and skewness on the daily basis. These two risk measures were estimated from the minute high-frequency data. The next we relax the assumption about no cross-sectional dependence and so we allowed the existence of the common factors influence all analyzed stocks. We found that both factors have the statistically significant impact on the return. The next the variance is negatively correlated with the return and skewness is positively correlated with the return. It means that the investors prefer volatility and negatively skew financial stocks which could imply the speculative behavior of the investors during the period 11.12.2007 to the 29.11.2016.

Keywords: Skewness, Cross-section dependence, APT, Common correlation effects

JEL classification: C33, G12
AMS classification: 91G70

1 Introduction

The main economics model for the understanding of asset pricing is Capital Asset Pricing Model (CAPM) from Sharpe (1964). The author assumes that under several conditions the risk premium $r_{it} - r_{ft}$ is given by

$$r_{it} = \beta_{0i} + \beta_1(r_{Mt} - r_{ft}) + \epsilon_{it} \tag{1}$$

where r_{it} is asset return i at time t, r_{ft} is return of risk free asset at time t, r_{Mt} is return of market portfolio Cochrane (2005) and ϵ_{it} is random variable meeting Gauss-Markov theorem. So in the equation (1) Sharpe assumes only one source of systematic risk - Market Risk. The concept of the single source of systematic risk could be the reason why the cross-asset variation in expected returns cannot be explained by the market beta (β_1) alone.

Fama and French came in 1995 with the three-factor model when they add SMB defined as the difference between the return on a portfolio of small size stocks and the return on a portfolio of large size stocks and HML defined as the difference between the return on a portfolio of high book-to-market value stocks and the return on a portfolio of low book-to-market value stocks. Another concept of risk we can find in the financial econometrics. For example, the Value at Risk model works with the standard deviation of returns (volatility). The risk concept of volatility, respectively variance used Li (2009) in his augmented CAPM model about asset variance estimated via GARCH model Bollerslev (1986). He found that variance is a statistically significant factor in the premium valuation and there exist the positive relationships between risk premium and variance. And so we get further to the relevancy of CAPM assumption about ϵ_{it}. We usually assume that the unconditional security return has normal or student distribution in GARCH model. As the point, for example, Harvey and Siddique (1999) assumption about normality is not real and the distribution is skew. Kraus and Litzenberger analyzed the effect of skewness in CAPM model yet in 1976. They add systematic skewness defined as unconditional co-skewness to the CAPM model and found that co-skewness has the negative impact on the risk premium. Harvey and Siddique (2000) continued in the Kraus and Litzenberger work and instead of unconditional co-skewness, they worked with conditional version. The authors argued that ceteris paribus, investors will prefer portfolios that are right-skewed to portfolios that are left-skewed. They reached the similar conclusion as Kraus and Litzenberger. They found that co–skewness can explain some parts of the nonsystematic components in cross-sectional variation and is negatively correlated with the risk premium.

Lets generalize the CAPM model to a k-factor model. We will assume that asset returns are generated by a linear factor model from Ross (1977)

$$r_{it} = \beta_{0i} + \beta_{1i}f_{1t} + \beta_{2i}f_{2t} + ... + \beta_{ki}f_{kt} + \epsilon_{it} \tag{2}$$

[1]Department of Econometrics, University of Economics in Prague, Winston Churchill Square 4, 13067 Prague, Czech Republic, xfryl00@vse.cz

ϵ_{it} is $i.i.d.$ with zero mean value, $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{kt})$ is $1 \times k$ vector of the factors and β are the factor loading. When no arbitrage opportunities exist than

$$E[r_{it} - r_{ft}] = \beta_{0i} + \beta_{1i} f_{1t} + \beta_{2i} f_{2t} + ... + \beta_{ki} f_{kt} \tag{3}$$

The problem is how many factors f exists and how many of them are visible. The next how real is assumtion $E[\epsilon_{it}\epsilon_{jt}] = 0$.

In this article we used the APT model with factors realized variance and skewness when we assume the existence of latent factors f which influence the regressors. In the first part, we will estimate the equation only with realized variance

$$r_{it} = \beta_0 + \beta_1 r_{it}^2 + \epsilon_{it} \tag{4}$$

and the next we add the skewness. We will analyze the importance of skewness in the stock pricing of the financial sector.

$$r_{it} = \beta_0 + \beta_1 r_{it}^2 + \beta_3 r_{it}^3 + \epsilon_{it} \tag{5}$$

definition of r^2, r^3 and assumtion about ϵ_{it} will be introduce in the next section.

2 Methodology

2.1 Realized moments

Following Amaya (2015) we assume that the log-price p_T evolves according to the stochastic equation:

$$p_T = \int_0^T \mu_s ds + \int_0^T \sigma_s dW_s + J_T \tag{6}$$

where μ is a drift process, σ is a positive càdlàg process, J is a pure jump process and W is a standard Brownian motion. The kth intraday return is defined as

$$r_{t,k} = p_{t,k/N} - p_{t,(k-1)/N} \tag{7}$$

when N is the number of return observations in a trading day.

Andersen, Bollerslev, Diebold, and Labys (2003) show that we can consistently estimate the quadratic variation as:

$$RM(2) = \sum_{i=1}^{N} (p_{T,\frac{k}{N}} - p_{T,\frac{k-1}{N}})^2 \xrightarrow{p} \int_0^T \sigma_s^2 ds + \sum_{0<s\leq T} (\Delta p_s)^2, as N \to \infty \tag{8}$$

where $\Delta p_s = p_s - p_{s-}$ is the size of the jump at time s.

The realized third moments are the consistent estimator of the part of cubic variation. As point Amaya (2015) RM(3) does not capture the part of skewness arising from the correlation between return and variance.

$$RM(3) = \sum_{i=1}^{N} (p_{T,\frac{k}{N}} - p_{T,\frac{k-1}{N}})^3 \xrightarrow{p} \sum_{0<s\leq T} (\Delta p_s)^3, as N \to \infty \tag{9}$$

2.2 Cross section dependence

We will assume the following APT model

$$r_{it} = \beta_{0i} + \beta_{1i} r_{it}^2 + \beta_{3i} r_{it}^3 + \epsilon_{it} \tag{10}$$

where $i = 1, ..., N,\ t = 1, ..., T$, ϵ_{it} is the composite error term.

From the reason of shocks we expecte cross–section dependence in the composite error ϵ_{it} and variables r_{it}^2 and r_{it}^3 . We adopted the data generation process of the composite error from Pesaran (2006) term following multi–factor structure

$$\epsilon_{it} = \boldsymbol{\gamma}_i^{'} \boldsymbol{f}_t + e_{it} \tag{11}$$

where $\boldsymbol{f}_t$ is the $m \times 1$ vector of m unobserved common factors (latent common factors) and the vector of parameters $\boldsymbol{\gamma}_i$ $m \times 1$ is known as the factor loading. The e_{it} is assumed to be $i.i.d.$ with zero expected value and independent on the regresors. The next we assume r_{it}^2 and r_{it}^3 have the following multi–factor structure:

$$r_{it}^2 = \boldsymbol{\theta_1}_i^{'} \boldsymbol{f}_t + w_{it} \tag{12}$$

$$r_{it}^3 = \boldsymbol{\theta_2}_i^{'} \boldsymbol{f}_t + u_{it} \tag{13}$$

where $\boldsymbol{\theta}_i$ is $m \times 1$ vector of factor loadings, w_{it} and u_{it} are the errors independently distributed on $\boldsymbol{f}_t$, e_{it} for all i, t. The next we assume the heterogeneous coefficients are randomly distributed around a common mean

$$\beta_i = \beta + \nu_i \tag{14}$$

when $\nu_i \sim iid(0, \Sigma)$, for $i = 1, 2, ..., N.$ and ν_i are distributed independently of e_{jt}, w_{jt}, $\boldsymbol{\theta}_j$ and $\boldsymbol{\gamma}_j$ for all i,j and t.

Estimation of (10) subject to restrictions (11), (12), (13) with Fixed effect estimator will lead to inconsistent estimations β Pesaran and Tosetti (2011), Moscone and Tosetti (2010). Pesaran (2006) suggest the estimator Common correlation effects (CCE) which provide the consistent estimation and valid inference of model represented by (10), (11), (12), (13). The main idea of CCE is based on the approximation of the latent common factors $\boldsymbol{f}_t$ by a linear combination of the cross section averages of dependent and independent variables.

We could augmented the equation about the averages of the cross-sectional variables od r, r^2, r^3.

$$r_{it} = \beta_{0i} + \beta_{1i} r_{it}^2 + \beta_{2i} r_{it}^3 + \delta_{1,i} \bar{r}_t + \delta_{2,i} \bar{r^2}_t + \delta_{3,i} \bar{r}_t^3 + v_{it} \tag{15}$$

where $\bar{r}_t = 1/N \sum_{i=1}^{N} r_{it}$, $\bar{r^2}_t = 1/N \sum_{i=1}^{N} r_{it}^2$ and $\bar{skew}_t = 1/N \sum_{i=1}^{N} r_{it}^3$ for all t. v_{it} is the error term consists of three parts. The error term e_{it} from the equation (11), error component due to the approximation of unobserved common factors and an error component due to the truncation of possibly infinite distributed lag function (Pesaran 2006).

The principle of CCEMG estimator is in the OLS application on the every i unit. We get the estimate of the β_i from (15) as the $\hat{\beta}_{CCE,i}$. Because we wish to identify common patterns of responses across otherwise heterogenous units the object of interest is $E(\beta_i) = \beta$. Then we apply

$$\hat{\beta}_{CCEMG} = N^{-1} \sum_{i=1}^{N} \hat{\beta}_{CCE,i} \tag{16}$$

Pesaran and Tosetti (2011) show that under some general condition for $(N, T) \to \infty$, CCE estimator is consistent and asymptotically normal:

$$\sqrt{N}(\hat{\beta}_{CCEMG} - \beta) \xrightarrow{d} N(0, \Sigma_{CCEMG}) \tag{17}$$

The consistent estimator of the Σ_{CCEMG} can be obtained by the non–parametric estimator, which is robust to autocorrelation and weak cross–sectional dependence in e_{it} .

3 Data

Our dataset contains minute data of 60 financial companies from the index $S\&P\ 500$. The time period span is from the 11.12.2007 to the 29.11.2016. We construct minute log-returns according to (7) with k equal to one minute and p_T is the log-price. Because our dataset contains gaps the N was not always equal. The daily realized moment estimator is the average of the available minute log-returns according (8) and (9). The summary statistics are displayed in the Table 1, where we can see the negative skewness of returns.

Variable	Min	Mean	Max	Variance	Skewness
r	-1.798e-03	-2.616e-07	1.199e-03	4.075e-09	-0.433
r^2	2.700e-08	4.001e-06	3.480e-03	4.601e-10	59.794
r^3	-8.41e-05	-2.05e-09	9.70e-05	4.224e-13	-11.493

Table 1 Summary statistics

4 Empirical part

In the first we test the cross–section dependence through the Pesaran (2004) test where the H_0 refers to the no cross-sectional dependence. The result is displayed in the Table 2. We can reject the H_0 on $\alpha = 0.01$.

CS	p-value
894.55	0.000

Table 2 Cross–section dependence test

In the next, we estimate the model (15) with the only one risk factor r^2. The results are displayed in the Table 3. We can see the statistical significant parameter β_1 for $\alpha = 0.01$. We do not display the estimation of the parameters δ because they are unnecessary.

Parameter	Coefficient	std. error	t-test	p-value
β_1	-0.206	0.068	-2.990	0.003

Table 3 Estimation of restricted equation (10)

The results for whole model (15) are displayed in the Table 4. We can reject the null hypothesis for both parameters on the $\alpha = 0.01$. The β_1 remains negative and very similar to the estimation from the model (4). The parameter β_2 is positive with estimation value 41.2. On the first it looks like a very high value but with respect to the summary statistics in the Table 1 it is economically reasonably number.

Parameter	Coefficient	std. error	t-test	p-value
β_1	-0.227	0.078	-2.88	0.004
β_2	41.216	5.366	7.68	0.000

Table 4 Estimation of equation (10)

5 Conclusion

In this article, we analyzed the significance of skewness on the return in the financial sector in the presence of cross–section dependence. We found that there exists the negative relationship between return and positive relationships between skewness. The interesting finding was the non-significant change in the variance premium when we added the skewness to the model. We expected that some premium value of variance would be transferred to the premium value of skewness. Next, we get the opposite signs of coefficient than Harvey and Siddique (2000), Li (2009). In our case, the higher variance means lower return which means that the investors prefer variance. On the other hand, the higher value of skewness increase the return and so investors prefer the negatively skew financial stocks. These results suggest the risky (speculative) behavior of the investors.

Acknowledgements

We gratefully acknowledge the support of this project by the research grant VSE IGA F4/78/2018, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Amaya, D.& Christoffersen, P. & Jacobs, K. & Vasquez, A.(2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118(1), 135–167.

[2] Andersen, T. G. & Bollerslev, T. & Diebold, F. X. & Labys, P. (1995). Modeling and forecasting realized volatility. *Econometrica*, 71, 529–626.

[3] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307–327.

[4] Cochrane, H. J. (2005). *Asset Pricing*. Princeton, New Jersey 08540: Princeton University Press.

[5] Fama, E. & French, K. (1995). Size and book-to-market factors in earnings and returns. *The Journal of Finance*, 50, 131–155.

[6] Harvey, C. R. & Siddique, A. (1999). Autoregressive Conditional Skewness. *The Journal of Financial and Quantitative Analysis*, 34(4), 465.

[7] Harvey, C. R. & Siddique, A. (2000). Conditional Skewness in Asset Pricing Tests. *The Journal of Finance*, 55(3), 1263–1295.

[8] Kraus, A. & Litzenberger, R. H. (1976). Skewness Preference and the Valuation of Risk Assets. *The Journal of Finance*, 31(4), 1085.

[9] Li, X.& Brooks, C.& Miffre, J. (2009). The Value Premium and Time-Varying Volatility. *Journal of Business Finance & Accounting*, 36(9–10), 1252–1272.

[10] Moscone, F. & Tosetti, E. (2010). Testing for error cross section independence with an application to US health expenditure. *Regional Science and Urban Economics*, 40(5), 283–291.

[11] Pesaran, M. H. (2004). General Diagnostic Tests for Cross Section Dependence in Panels. *CESifo Group Munich CESifo Working Paper Series 1229*

[12] Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with multifactor error structure. *Econometrica*, 74, 967–1012.

[13] Pesaran, M. H. & Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2), 182–202.

[14] Ross, S. A. (1977). The Capital Asset Pricing Model (Capm), Short-Sale Restrictions And Related Issues. *The Journal of Finance*, 32(1), 177–183.

[15] Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19, 425–442.

Calculation of regional disparities in preventable mortality in Slovak regions

Beáta Gavurová[1], Peter Tóth[2]

Abstract. One of the key issues in the health care management are the preventive programmes. Preventable mortality is a significant part of the mortality, which can be eliminated by an appropriate prevention programme. There are several approaches to prevention, such as healthy lifestyle or healthy environment. There are the studies which analyse the impact of urban areas, districts and geographical areas on the preventable mortality. The main aim of the paper is to analyse preventable mortality in the regions of the Slovak Republic and to identify the causes of deaths with the highest contributions to the overall regional mortality. To calculate preventable mortality, a direct death standardisation method is applied. The regional disparities obtained from the analysis can provide a base for prevention programming, which takes into account regional needs. This can lead to decrease in preventable mortality in general as well as reduction of the regional disparities in preventable mortality.

Keywords: preventable mortality, regional disparities, health care system, standardised mortality rate.

JEL classification: C46, I10
AMS classification: 91B82

1 Introduction

Ageing of the population is determined primarily by the average life expectancy, the economics conditions in society, the level of health care, family environment, environmental quality, individual lifestyle, education, etc [10]. Demographic processes are not homogeneous in the regions of the Slovak republic, they differ in speed of change and their intensity, which increases differences in demographic aging [4]. Health improvement has been traditionally measured by the mortality rates, life expectancy, premature mortality or by the so-called 'Lost years of life'. The choice of these metrics has been determined primarily by the data availability as well as possibility of measuring and comparisons in time horizons at national and international level [11]. These measurements have not reflected the disease burden, which can be attributed to the health system. The aim of each health system should be effective support, restoration and maintenance of the population health. In order to quantify and evaluate them, it is important to explore many issues related to the functioning of the health system [12]. One of them is avoidable mortality, which can provide specific and more complex view about health care and the level of health interventions.

According to the [9], avoidable mortality consists of two groups: amenable and preventable deaths. Amenable mortality is the number of deaths that, based on the medical knowledge and technologies available at the time of death, all or most of deaths caused by the specified diseases should not occur in case of providing adequate health care. Preventable mortality denotes the number of deaths, based on the medical knowledge and technologies available at the time of death, all or most of deaths caused by the specified diseases should not occur in case of effective public health campaigns aimed at prevention.

Many research studies deal with the development of the avoidable mortality and determinants related to it. E.g. [5] presented in their study maps of the avoidable mortality in European cities and analysed differences in avoidable mortality between adjacent areas with different levels of deprivation. The results of their studies confirm the existence of significant differences in the level of avoidable mortality between adjacent regions of European cities. They confirm that the rate of avoidable mortality is associated with social deprivation. [6] examined the relationship between income and inequalities in avoidable mortality - amenable and preventable in Norway. They observed decreasing trend in avoidable, amenable and preventable mortality in Norway. According to them, income-related inequalities in avoidable, amenable and preventable mortality remained relatively constant between 1994 and 2011 in Norway.

The quality and performance of the Slovak and Czech health systems before 1989 (1971-1989) and after 1989 (1990-2008) is examined by [7]. Her analysis includes the year of division of the republic in 1993 using the concept of avoidable mortality. The results of their analyses show that post-communist healthcare reforms and

[1]Technical University of Košice, Faculty of Economics, Němcovej 32, 04001 Košice, beata.gavurova@tuke.sk
[2]Technical University of Košice, Faculty of Economics, Němcovej 32, 04001 Košice, peter.toth@tuke.sk

the disintegration of Czechoslovakia have been explored from a different perspective, but the performance of healthcare systems has been under-researched. At the time of partition, the republic of Slovakia slightly lagged behind the Czech Republic, mainly due to higher mortality rates for ischemic heart disease and hypertension and cerebrovascular diseases. This gap can be explained by socio-economic, environmental differences as well as differences in living standards between countries. In the following years the differences between the countries gradually eroded, in some indicators Slovakia even achieved better results than the Czech Republic. The decline in overall reversible mortality suggests improving the performance of the healthcare system and the quality of healthcare in both countries. [13] analysed the relationship of education inequality to the development of avoidable mortality in 16 European countries. Their ambition was to prepare an overview of a wide range of possible causes of death as well as to determine the contribution of disparities in avoidable mortality to inequality among the most educated and least educated population. The results show that differences in education levels have also been reflected in the results of avoidable mortality in all European countries. Inequalities in avoidable mortality were present in all European countries, but were particularly significant in the countries of Central and Eastern Europe and the Baltic countries. These educational inequalities point to the important role of health care services in reducing health inequalities.

In [8], the authors outlined the potential reasons for the relationship between avoidable mortality and socio-economic inequality. They explain that social inequalities in health arise due to inequalities in society such as the conditions of everyday life, education, money and resources. This means that people with a higher social status are more likely to have a better health than people with a lower social status, and thus a lower risk of premature death. Although the results of research studies, given the focus of research teams and study objectives, are very heterogeneous, they provide many valuable information for state policy makers and for national and international benchmarking. By comparing the regions within countries with results obtained in other countries, it will help identify the necessary primary and secondary prevention programs for the region and thus increase the efficiency of healthcare provision. These facts were the motivation for our research analyses focused on the situation in the field of preventable mortality in Slovakia, of which the partial section is presented in the following subchapters.

2 Data and Methodology

The paper analysis mortality data in Slovak republic provided by the National Health Information Centre (Národné centrum zdravotníckych informácií) of the Slovak republic. Dataset consists of all deceased since 1996. In this paper we analyse preventable mortality in 2015, which is the last available year in the dataset.

The concept of avoidable mortality is based on the idea that deaths caused by certain purposes should be scarce or should not occur when suitable prevention programmes and effective public health system exist. List of causes of preventable mortality is set by the [9] and accepted by the European Commission. Complete list is presented in Table 1. Preventable mortality causes are divided into eight groups. The first group consists of Infections, second group includes Neoplasms, next are Nutritional, endocrine and metabolic diseases, then Drug use disorders, Neurological disorders, Cardiovascular diseases, Respiratory disorders, Unintentional injuries and the last group covers Intentional injuries.

Calculation of the preventable mortality in district i is given as a sum of standardized death rates (SDR) for specified causes and age categories. According to [2] and [1], SDR is calculated by equation (1), where x denotes age category $0, 1-4, 5-9, \ldots, 90-95, 95+$, m_{ix} is age-specific death rate and ESP represents European Standard Population established by [3].

$$SDR_i = \frac{\sum_x m_{ix} ESP_x}{\sum_x ESP_x} \, 100\,000 \tag{1}$$

Age-specific death rate m_{ix} is expressed by equation (2), where D_x denotes number of deceased in age category x in district i and P_{ix} is average population in age category x in district i.

$$m_{ix} = \frac{D_{ix}}{P_{ix}} \tag{2}$$

3 Results

Analysis is divided into two parts. In the first part, we analyse regional differences in preventable mortality in Slovak regions. In the second part, we study causes of death with the highest contribution to the preventable mortality in Slovak regions.

As well as many social and economic indicators, preventable mortality is very heterogeneous among Slovak districts. In 2015, the highest preventable mortality for men was 719.86 in district Rimavská Sobota. The second

highest preventable mortality was 711.39 in district Veľký Krtíš and the third was 693.83 in district Žarnovica. Districts Rimavská Sobota and Veľký Krtíš belong to the least developed districts in the Slovak republic. On the other hand, the lowest preventable mortality was 283.73 in district Bratislava I, 307.61 in district Bratislava IV and 311.67 in district Košice I. all these three districts are districts in the largest cities in the Slovak republic. The average value of the preventable mortality for men in 2015 was 485.02.

Table 1 Preventable mortality cause list

Cause	ICD-10 codes	Age
Tuberculosis	A15-A19, B90	0-74
Hepatitis C	B17.1, B-18.2	0-74
HIV/AIDS	B20-B24	all
Malignant neoplasm of lip, oral cavity and pharynx	C00-C14	0-74
Malignant neoplasm of oesophagus	C15	0-74
Malignant neoplasm of stomach	C16	0-74
Malignant neoplasm of colon and rectum	C18-C21	0-74
Malignant neoplasm of liver	C22	0-74
Malignant neoplasm of trachea, bronchus and lung	C33-C34	0-74
Malignant melanoma of skin	C43	0-74
Mesothelioma	C45	0-74
Malignant neoplasm of breast	C50	0-74
Malignant neoplasm of cervix uteri	C53	0-74
Diabetes mellitus	E10-E14	0-49
Alcohol related diseases, excluding external causes	F10, G31.2,G62.1, I42.6, K29.2, K70, K73, K74 (excl. K74.3-K74.5), K86.0	0-74
Illicit drug use disorders	F11-F16, F18-F19	0-74
Ischaemic heart diseases	I20-I25	0-74
DVT with pulmonary embolism	I26, I80.1-I80.3, I80.9, I82.9	0-74
Aortic aneurysm and dissection	I71	0-74
Influenza incl. swine flu	J09 - J11	0-74
Chronic obstructive pulmonary disorder	J40-J44	0-74
Transport accidents	V01-V99	all
Accidental injury	W00-X59	all
Suicide and self inflicted injuries	X60-X84, Y10-Y34	all
Homicide/Assault	X85-Y09, U50.9	all
Misadventures to patients during surgical and medical care	Y60-Y69, Y83-Y84	all

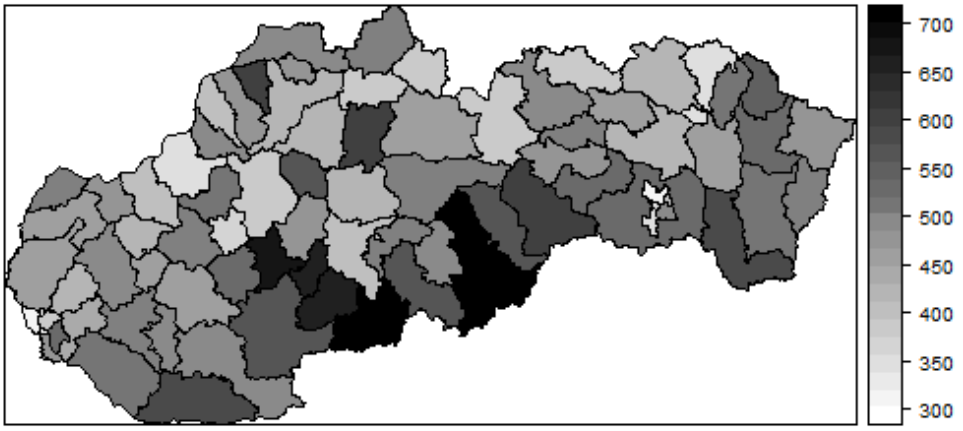


Figure 1 Preventable mortality in Slovak districts for men in 2015

In case of women, the average value of the preventable mortality was 192.08, which is less than half of the value for men. The highest preventable mortality was 273.57, again in district Rimavská Sobota, followed by the district Krupina with 265.76 and district Revúca with preventable mortality 263.57. The lowest value of the preventable mortality is 108.73 in district Trenčín, then 123.22 in district Stará Ľubovňa and 128.17 in district Detva.

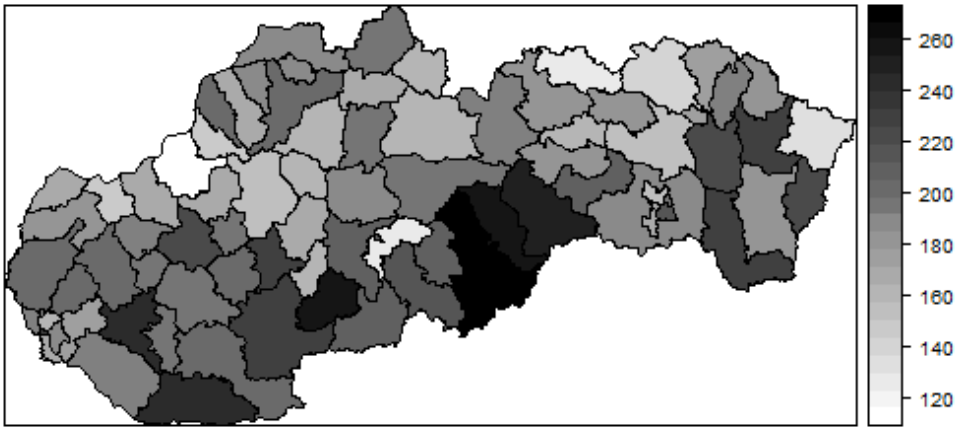


Figure 2 Preventable mortality in Slovak districts for women in 2015

Figure 1 shows the preventable mortality for men in Slovak districts in 2015. In general, we can conclude that the highest preventable mortality is in southern districts and the lowest is in the northern part of the Slovak republic. There are some districts, which do not fulfil that rule, e.g. Bytča and Ružomberok located in the North with high preventable mortality as well as districts like Stropkov, Medzilaborce and Humenné situated on the north-east. In two largest cities, which are divided into several districts (Bratislava - 5 districts and Košice - 4 districts), there are significant differences among districts in the city. While districts Bratislava I and Bratislava IV belongs to the districts with the lowest preventable mortality, district Bratislava II is above the average. The same situation is in Košice city, where districts Košice I and Košice II are the third and the fourth best districts, respectively and district Košice IV is district with the average level of the preventable mortality in case of men.

Table 2 The most frequent preventable causes in Slovak districts in 2015

	Men		Women	
	Diagnosis	Score	Diagnosis	Score
1	Ischaemic heart disease	80	Ischaemic heart disease	93
2	Malignant neoplasm of trachea, bronchus and lung	250	Malignant neoplasm of breast	276
3	Suicide and self inflicted injuries	347	Suicide and self inflicted injuries	438
4	Alcohol related diseases, excluding external causes	356	Malignant neoplasm of colon and rectum	448
5	Malignant neoplasm of colon and rectum	437	Malignant neoplasm of trachea, bronchus and lung	465
6	Pneumonia	494	Alcohol related diseases, excluding external causes	474
7	Malignant neoplasm of lip, oral cavity and pharynx	623	Pneumonia	543
8	Accidental Injury	638	Accidental Injury	754
9	Transport Accidents	803	Malignant neoplasm of cervix uteri	760
10	Malignant neoplasm of stomach	833	Malignant neoplasm of stomach	805

Similar geographical layout is in the case of women, with districts with the lowest preventable mortality in the northern part and districts with the highest mortality in the southern part. There are several exceptions, such as districts Detva and Banská Štiavnica located in the southern part of the Slovak republic with the low preventable mortality. Same as in case of men, districts on the north-east part of the country such as Vranov nad Topľou and

Humenné have high preventable mortality. There are not significant differences among districts in two largest cities such as in case of men.

The second aim of the paper is to identify the most frequent causes of death, which are included in to the preventable mortality. We applied the score method, where score is calculated as the sum of orders from all districts. For both sexes, Ischaemic hearth disease is the cause with the highest contribution to the preventable mortality. The score for men is 80, which means that according to the standardized death rate, in 78 districts was on the first place and in one district was on the second place. For men, the second cause is Malignant neoplasm of trachea, bronchus and lung, then follow Suicide and self inflicted injuries. The fourth most common diagnosis are Alcohol related diseases. In case of women, the second cause is Malignant neoplasm of breast and the third is Suicide and self inflicted injuries followed by the Malignant neoplasm of colon and rectum. More detailed results are presented in Table 2. If we look at the score difference between the tenth cause and the first cause, we can see that this difference is higher in case of men, which means that the order of causes of death in districts is more similar than in the case of women.

4 Conclusion

Aim of the paper was to analyse preventable mortality in Slovak regions and identify most frequent causes of preventable mortality. Issue of preventable mortality is crucial in process of prevention programming. In order to set suitable prevention program, identification of main causes of preventable mortality is needed. Slovak republic is heterogeneous from different perspectives. Our analysis confirmed that it is true for preventable mortality, too. Regions differ in the level of preventable mortality. To generalize, we can claim that high preventable mortality is in the south part of the country and low is in districts located in the north. Special issue in Slovak republic are least developed regions selected according to the unemployment rate. These regions are situated in the south part and east part of the country. Actually, there are 16 districts considered as least developed regions. To sum up, in these regions is preventable above average, especially these in the south. Heterogeneity is typical also for large cities, such as Bratislava and Košice, where exist districts with high preventable mortality next to districts with low preventable mortality. That research provides base analysis of preventable mortality in Slovak regions. To prepare appropriate prevention program, further analysis is needed. It is important to identify differences among regions in causes of preventable mortality. In this study, we analysed most frequent causes of deaths in districts. The most frequent cause of death in case of both sexes was Ischaemic heart disease followed by Malignant neoplasm of trachea, bronchus and lung in case of men and Malignant neoplasm of breast in case of women.

Acknowledgements

This work wassupported by the VEGA Project No. 1/0945/17 Economic research on quantification of marketing processes aimed at improving value for patient, multidimensional analyses of the marketing mix of healthcare facilities and quantification of their importance in the process of establishment of the system to measure the quality and efficiency in healthcare of the Slovak Republic.

References

[1] Anderson, R. N., & Rosenberg, H. M. (1998). Age standardization of death rates: implementation of the year 2000 standard. National vital statistics reports, 47(3), 1-17.

[2] Curtin, L. R., & Klein, R. J. (1995). Direct standardization (age-adjusted death rates) (No. 6). US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Health Statistics.

[3] European Commission (2013). Revision of the European Standard Population – Report of Eurostat‘s task force. Luxembourg: Publication Office of the European Union.

[4] Gavurová, B., & Vagašová, T. (2017). Assessment of Avoidable Mortality Concepts in the European Union Countries, Their Benefits and Limitations. In Advances in Health Management. InTech. http://dx.doi.org/10.5772/67818

[5] Hoffmann, R., Borsboom, G., Saez, M., DellOlmo, M. M., Burstrm, B., Corman, D., ... & Gandarillas, A. (2014). Social differences in avoidable mortality between small areas of 15 European cities: an ecological study. International Journal of Health Geographics, 13(1), 8. doi:10.1186/1476-072X-13-8.

[6] Kinge, J. M., Vallejo-Torres, L., & Morris, S. (2015). Income related inequalities in avoidable mortality in Norway: A population-based study using data from 19942011. Health Policy, 119(7), 889-898.

[7] Kossarova, L., Holland, W., & Mossialos, E. (2013). Avoidablemortality: a measure of health system performance in the Czech Republic and Slovakia between 1971 and 2008. Health policy and planning, 28(5), 508-525. doi:10.1093/heapol/czs093

[8] Marmot, M., & Bell, R. (2012). Fair society, healthy lives. Public health, 126, S4-S10.

[9] Office for National Statistics. 2013. Avoidable mortality in England and Wales, 2011. Experimental statistics. [online] [cited 2013-10-09]. Available from Internet: www.ons.gov.uk

[10] Plug, I., Hoffman, R., Mackenbach, J. (2011). AMIEHS. Avoidable mortality in the European Union: Towards better indicators for the effectiveness of health systems. EU Public Health Program 2007206.

[11] Pol, L., Thomas, R.K. (2013). The Demography of Health and Healthcare. (Third Edition). London: Springer, Springer Dordrecht Heidelberg New York London, 2013, 305 p. ISBN 978-90-481-8902-1.

[12] Štefko, R., Jencova, S., Litavcova, E. (2017). Management and funding of the healthcare system. Polish Journal of Management Studies, (16)2, pp. 266-277.

[13] Stirbu, I., Kunst, A. E., Bopp, M., Leinsalu, M., Regidor, E., Esnaola, S., ... & Kalediene, R. (2010). Educational inequalities in avoidable mortality in Europe. Journal of Epidemiology & Community Health, 64(10), 913-920. doi:10.1136/jech.2008.081737

Factors Influencing TV Advertising Effectiveness: Bayesian Networks Application

Daria Gunina[1], Lucie Váchová[2]

Abstract. Advertising management involves decision-making concerning media scheduling, i.e. with the aim to increase advertising effectiveness. TV commercials are aimed to reach a broad audience, so the more people are exposed to the message, the more expensive and the more effective the commercial is expected to be. In this paper, advertising effectiveness is construed as the maximisation of advertising exposure. In previous studies, several statistically significant features influencing advertising effectiveness were revealed; some of them showed a small impact and some of them indicated multicollinearity. Hence, the aim of our research is to infer the relationships between the features that can influence advertising effectiveness using Bayesian network modelling. The research is based on complete monitoring data on advertising on the Czech TV market in 2017.

Keywords: advertising exposure, TV commercials, Bayesian network.

JEL classification: M37, C11
AMS classification: 90B60, 62C12

1 Introduction

The effectiveness of television advertising has been the main goal of advertisers for a long time [18], [20] and still remains the most important issue in the modern era of digitalisation and media proliferation [6]. Advertising effectiveness (AE) has many different definitions and can be differently measured. Some researchers in the field of advertising attempted to classify its concepts into an "advertising effect pyramid" [13]. This pyramid includes several levels representing different measures of AE: exposure, awareness, comprehension, recall, conviction, desire, and action. Exposure provides information on how many viewers were exposed to the advertising message (watched the whole spot, or watched at least three minutes if a spot is longer than three minutes), i.e. the share of the population that was exposed. Exposure is at the bottom of the pyramid: people are exposed every day to countless advertisements, so each ad they receive is effective (if its exposure is measured). Action (or purchase) is at the top of the pyramid, because only small share of advertisements can be considered to be effective according to this measure. Yet the relationship between advertising activity and sales remains unclear. For example, advertising activity has a positive effect on sales on the car market [12] and cigarette market [5], but not on the beer [15], whisky [4] or coffee market [5]. Moreover, companies may set other advertising goals than increasing sales: i.e. improving brand image or brand awareness.

Exposure is underestimated by advertisers because their advertising campaigns are not usually aimed at maximising exposure, but rather on consumer recall or purchases (action). However, exposure should be considered as important, as it indicates the number (or share) of ads that *CAN* become effective on the next level of the pyramid and have the potential to result in a purchase decision. Ads that are not effective on the bottom are excluded from the "AE competition" - unless the ad is viewed, it cannot influence purchase decision-making. "Exposure level" acts as a filter: if an ad is not effective in the meaning of exposure, it will not be effective by the other meanings of AE. Moreover, exposure is the only AE meaning that depends solely on media scheduling, and contrary to other "indicators" (recall, liking, sells etc.) does not include the influence of advertising content and creativity. Moreover, the number of TV viewers is declining and it is harder to address a mass audience [6]; therefore, measuring advertising exposure is one way of monitoring this trend.

Advertising media scheduling solves questions such as what to air, where to air, and when to air [14]. Media scheduling can differ for different *Product categories*, since high-involvement products (i.e. cars) and low-involvement products (i.e. drugstore goods) require a different media communication strategy [11]. The *Group of channels* and *Channel* variables explain where to air. The question "when to air" is answered by such media scheduling variables as the *Month of airing*, *Week of airing*, *Day of airing* and *Time of airing*. These variables are adapted to seasonality and audience lifestyle, workweek and weekend, working and sleeping hours. Katz [11] states that media space selling/buying has two criterions in the USA: *Daypart* (nine different parts of the day) and format,

[1]University of Economics, Prague/ Faculty of Management, Department of Management, Jarošovská 1117/II, Jindřichův Hradec, Czech Republic, daria.gunina@vse.cz
[2]University of Economics, Prague/ Faculty of Management, Department of Exact Methods, Jarošovská 1117/II, Jindřichův Hradec, Czech Republic, lucie.vachova@vse.cz

which means the *Programme type* during when the ad is aired (twelve different types). Media context/environment (particularly the *Programme type* on TV) can radically influence ad exposure and other effects [21]. Various scholars [2], [13] have analysed the influence of *Length of a spot*, *Number of spots in a break* and *Position in a break*. Too short commercials make "clutter" effect and are considered to have a deleterious influence on AE [5]. In his research, Danaher [7] found that *Position in a break*, *Number of ads in a break*, *Length of a break* and *Length of a spot* influence exposure, but not considerably; however, the *Programme type* has a strong influence. However, the level of exposure cannot be fully explained solely by these media scheduling indicators [7].

Hence, the aim of our research is to examine the probability relationships among the media features using Bayesian network (BN) modelling.

2 Data and Methodology

Since their origin, Bayesian networks have been used in a broad range of application areas, especially as a tool for representing probabilistic knowledge. They are also often used in management and marketing research [1], [3], [8], [19]. In a Bayesian network, the variables of interest are represented by the nodes while the informational or causal dependencies among these variables are represented by the links [16]. Given that $V = \{V_1, ..., V_n\}$ is a set of variables in the model, and v_i denote the value of the variable V_i we get

$$P(v_1, ..., v_n) \quad = \prod_i P(v_i|pa_i) \tag{1}$$

where pa_i is a set of values for parents of variable V_i. For details on BN see, e.g., [10].

The motivation for deploying a Bayesian network for the TV advertising analysis stems from several aspects:

- the knowledge of advertising effectiveness is not complete;
- the knowledge of advertising effectiveness is uncertain;
- the suggestion is that the advertising effectiveness field is of a probabilistic nature;
- BN and related software enables the research to be built on real data with a huge number of observations;
- BN can be deployed in research for descriptive, inference and predictive purpose;
- the design and conditional probabilities of BN constructing facilitate results explanation and presentation to a wide academic and practitioner public.

This study is based on the secondary dataset provided by the Nielsen Admosphere research agency. The sample contains data for all commercials aired on television in the Czech Republic in 2017 and includes more than 5,608 million observations. The observation in this study is not the commercial itself but each airing of a particular commercial, as one commercial could be aired several times and each airing could result in different levels of effectiveness. In the original dataset, each observation was described by 14 variables *Group of channels* (TV outlet), *TV channel*, *Programme type (before a break)*, *Programme type (after a break)*, *Length of a break*, *Length of a spot*, *Number of spots in a break*, *Position of a spot in a break*, *Month of airing*, *Week of airing*, *Day of airing*, *Time of airing*, *Product category* and *Exposure*. For the purpose of this paper, the dataset was slightly modified and the new set of 13 variables used in the analysis is shown in Table 1.

Variable	Label	Number of values	Variable	Label	Number of values
Group of channels	GC	7	TV Channel	CH	45
Programme type (before a break)	PTB	57	Programme type (after a break)	PTA	57
Length of a break	LB	24	Length of a spot	LS	13
Number of spots in a break	NS	30	Position of a spot in a break	PS	4
Week of airing	WA	53	Weekday of airing	DA	7
Daypart	PD	7	Product category	PC	22
Exposure	ES	6			

Table 1 Input variables

The *Length of a break* and *Length of a spot* variables were transformed into intervals based on the lengths usually used in TV advertising literature. The *Month of airing* was omitted because in the previous study multicollinearity was proven between the *Month of airing* and *Week of airing*. The *Time of airing* (hour-minute-second) was transformed into the form of the *Daypart*. Katz [11] defined nine dayparts for the American market. As the Czech media market differs from the American market, the dayparts were defined according to media habits in the

Czech Republic: Late night (11 p.m. - 1 a.m.), Dead time (1 - 6 a.m.), Early morning (6 - 9 a.m.), Daytime (9 a.m. - 12 p.m.), Early fringe (12 - 5 p.m.), Evening (5 - 7 p.m.) and Primetime (7 - 11 p.m.). As a product category, the NACE classification on the first level was used, which in our case included 22 categories. *Exposure* (originally a continuous variable) was also expressed in the form of intervals with respect to the ability to analyse the generally interesting intervals (as set in TV advertising research). Only the observations with a complete description were used, resulting in the final dataset of 5 488 866 observations.

A quantitative approach was employed in this study in accordance with the quasi-experimental nature of research based on real data. The quantitative research design approach is recommended for a large number of observations in a sample. Data was analysed using the R software [17]. For the construction of the Bayesian network, the bnlearn package was used, which contains several algorithms for BN structure learning.

3 Results

To build a BN in our research, the Hill-Climbing algorithm was used. It assigns the score to each BN candidate and is based on the maximisation of this score. Each node in the resulting BN (as presented in Figure 1) indicates one variable from Table 1. Since *Exposure* is a variable attesting (economic) TV advertising effectiveness, it is an important finding that four variables have a direct relation with *Exposure* in the created BN: *TV channel*, *Weekday of airing*, *Week of airing* and *Daypart*. For the purpose of this study, we define "meaningful exposure" as exposure to 1% of the population or more; 469585 observations met this condition in our dataset.

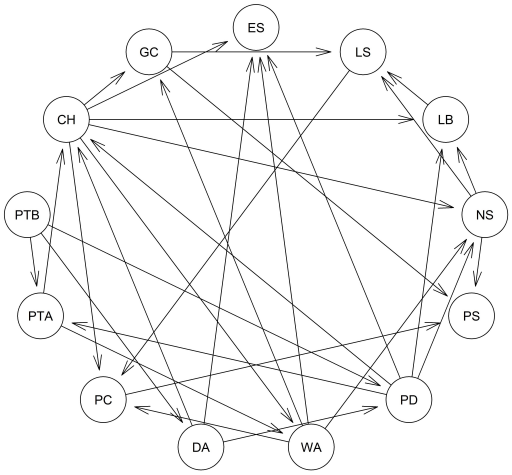


Figure 1 Bayesian network of media features

It is expected that during Primetime (7 - 11 p.m.), ads are more likely to have meaningful exposure than during the other *Dayparts*. During the Dead time (1 - 6 a.m.), the exposure is expected to be low. The results do not fully confirm these suggestions. The probability of meaningful exposure, given the Dead time as a value of *Daypart*, was 0.00% for all values of the *Weekday of airing*. However, Primetime did not always show the highest probability of meaningful exposure. On Mondays, Wednesdays and Sundays, the probability of meaningful exposure, given the *Daypart* and *Weekday of airing*, is higher in the evenings (5 - 7 p.m.) than during Primetime (see Table 2).

	Primetime	Evening
Monday	14.39	15.77
Tuesday	16.24	14.00
Wednesday	13.73	16.53
Thursday	18.53	11.87
Friday	16.26	15.77
Saturday	15.36	12.58
Sunday	16.45	16.99

Table 2 Probability of meaningful exposure given *Daypart* and *Weekday of airing* (in %)

These findings differ in different *Weeks of airing* (see Table 3; the 51st and 52nd weeks are examples). In this case, we studied the probability of meaningful exposure given the *Daypart*, *Weekday of Airing* and *Week of Airing*.

The results mostly show the different values in the 51st and 52nd weeks. For example, considering it is Wednesday and Daytime, it can be observed that the given probability in the 52nd week is 0.00%, in comparison with 13.3% in the 51st week. While an ad in the Evening on Sunday has a 10% probability of meaningful exposure in the 51st week, in the 52nd week the probability is 62.5%.

		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Daytime	51st Week	10.00	16.67	13.33	07.69	06.25	00.00	12.50
Daytime	52nd Week	11.11	33.33	00.00	00.00	09.09	00.00	09.09
Early fringe	51st Week	10.34	14.29	14.29	11.11	07.14	03.70	13.33
Early fringe	52nd Week	00.00	08.33	00.00	09.09	07.69	18.18	16.67
Evening	51st Week	16.67	00.00	07.14	12.50	23.08	20.00	10.00
Evening	52nd Week	00.00	25.00	00.00	00.00	22.22	00.00	62.50
Primetime	51st Week	37.50	11.76	16.67	20.00	07.69	08.70	12.50
Primetime	52nd Week	00.00	17.65	00.00	14.29	20.00	15.00	33.33
Late night	51st Week	00.00	00.00	00.00	00.00	15.38	18.18	00.00
Late night	52nd Week	14.29	00.00	00.00	00.00	20.00	00.00	00.00

Table 3 Probability of meaningful exposure given *Daypart*, *Weekday of airing* and *Week of airing* (in %)

According to the literature review, the most frequently used *Lengths of spot* are 10, 15, 30 and 45 seconds. However, the results show that these lengths are not the most likely to have meaningful exposure. A 45-second ad has a 15% probability of being exposed to at least 1% of the population; in the case of 30, 15, and 10-second ads, the probabilities are 7%, 8.2%, and 13.8%, respectively. Surprisingly, for the shortest ads (with a length less than or equal to 9 seconds), the considered probability is the highest (18.5%). Additionally, the long commercials (with the length equal to or higher than 60 seconds) are not appropriate as they have the probability to achieve meaningful exposure lower than 2%.

The probability of achieving meaningful exposure can also be analysed with respect to the *Length of a break* and the *Number of spots in a break*. Considering the *Length of a break*, a non-typical finding is that the shortest and the longest breaks are more likely to have meaningful exposure (with probabilities of 20.3% and 19.4%). While given that the length of the ads is between 3:50 and 4:00 minutes, the probability is 1.8%, and for the more frequently used (101,919 observations) lengths between 5:58 and 6:00 minutes, the probability is only 1.2%. When it comes to *Number of spots in a break*, it was found that a larger number of spots in a break usually lead to a lower probability of meaningful exposure: for 24 or more spots in a break, the probability is 0.00%. However, setting the number of spots in a break from the interval between 20 and 23 seems to be more suitable, considering the probabilities of meaningful exposure as follows: for 20 spots, the probability is 24.6%; for 21 spots 26%; for 22 spots 20%; and for 23 spots 22.2%. We expected that the lower the number of spots, the higher the probability of meaningful exposure. This expectation was not confirmed: for a break with only 1 spot, the probability was 17.3%; with 2 spots 17.9%; and with 3 spots 13%.

According to previous studies (e.g., [7], [13]), the first and the last ad in a break are more likely to have higher exposure. We disprove this assumption about the *Position in a break*. The first, middle and last ads have similar probabilities of meaningful exposure, namely, 8.3% for the first, 8.4% for middle and 7.8% for the last ad in a break (regardless of the number of spots in a break). Considering that there are only 3 spots in a break, the probability of achieving the meaningful exposure is highest for the middle spot (17.6%), in comparison with 10.3% for the first ad and 12.3% for the last. In the case that the break consists of only one spot, we considered this as a special value, since such an ad would be simultaneously in both the first and the last positions. Therefore, the probability of meaningful exposure given the *Position in a break* being "the only one" is the same where the *Number of spots in a break* equals 1, thus 17.3%.

The next case analysed was the probability of meaningful exposure, given the *Product category* and the *Programme type* (before or after a break). In the Table 4, only those results are shown that include the probabilities of considered exposure, given the *Product category* and the *Programme type before a break* that were higher than 30%. Advertisers in different product categories use different programme types, as found from the frequencies of spots broadcasted in particular categories. Different programmes are watched by different target groups with various preferences, lifestyles, and needs. Thus, it can be assumed that the same ad broadcast in different programmes can differ in success with a rate of exposure. Hence, the aim of advertisers and media planners should be to match the product category (with the assumed target group for the promoted product) with the programme type (considering the related target audience for a programme). It can be observed in the Bayesian network created based on the data (see Figure 1), that the *Product category* and the *Programme type* (before or after a break) do not

have a direct relationship.

The limitation of this study stems from the limitation of the Bayesian networks construction. The structure of the BN was created from data with the use of one algorithm, and relationships in the resulting network were compared with the usual assumptions about the existence of such connections among the variables under consideration. No arrow was deleted, and none was added. In the analysis, 13 available variables were used to construct the Bayesian network and examine the probabilities of meaningful exposure. However, in reality, many additional factors may exist that are not currently measured but still influence advertising exposure. Advertising exposure is the measure of how many PEOPLE watched the ad. As soon as this variable is also influenced by consumer behaviour (with related problematic topics such as irrational decision-making, the human factor, consumer black box etc.), the model cannot fully reflect the real behaviour of consumers. However, it can still be considered as a useful tool for at least obtaining a preliminary image of the TV ad viewers' habits (what is watched and what is not). Another limitation strongly relates to data available for the analysis. The dataset contains some missing values, and some combinations of the values of the variables do not appear in the dataset, thus the exposure of such situations cannot be studied.

4 Conclusion

In this study we examined the probability relationships between the media features using Bayesian network modelling, which enabled us to compute conditional probabilities of "meaningful exposure", given the values of the selected set of other variables. The variables that have a direct relation with *Exposure* are *TV channel*, *Weekday of airing*, *Week of airing* and *Daypart*. We examined the probabilities of meaningful exposure in several situations: a) given the *Product category* and *Type of programme before a break* (an example of the results is shown in Table 4), b) given the *Daypart*, *Weekday of airing* and *Week of airing* (see Table 3 for results from the 51st and 52nd week of 2017), c) given the *Daypart* and *Weekday of airing* (see Table 2), d) given the *Length of spot*, e) given the *Length of break*, f) given the *Number of spots in a break*, g) given the *Position of a spot in a break*, and h) given the *Number of spots in a break* and the *Position of a spot in a break*.

Advertisers face the problem of a lack of information, as it is not known beforehand how many people will be exposed to a particular ad. We show that a Bayesian network can be considered to be a suitable tool for media planning. Based on historical data, it provides useful and important information and can be used in future decision-making. The model shows the relations among the variables of interest and also enables to compute the probabilities of a particular exposure, given the selected set of conditions.

Acknowledgements

This work was supported by Grant No. F6/01/2018 from the Internal Grant Agency FM VŠE. The authors would like to thank Ing. Michal Novák, Ph.D., University of Economics, Prague, for his helpful suggestions and comments.

References

[1] Anderson, R. D., and Lenz, R. T. (2001). Modeling the impact of organizational change: A Bayesian network approach. *Organizational Research Methods*, 4(2), 112-130.

[2] Billett, J. (1993). *Share of Break Study*. London: The Billett Consultancy.

[3] Blodgett, J. G., and Anderson, R. D. (2000). A Bayesian network model of the consumer complaint process. *Journal of Service Research*, 2(4), 321-338.

[4] Cavaliere, G. and Tassinari, G. (2001). Advertising effect on primary demand: a cointegration approach. *International Journal of Advertising*, 20(3), 319-339.

[5] Cowling, K., Cable, J., Kelly, M. and McGuiness, T. (1975). *Advertising and Economic Behaviour*. London: The Macmillan Press.

[6] Danaher, P. J. (2017). Advertising Effectiveness and Media Exposure. In *Handbook of Marketing Decision Models* (pp. 463-481). Springer.

[7] Danaher, P. J. (1995). What Happens to Television Ratings During Television Commercials? *Journal of Advertising Research*, 35(1), 37-47.

[8] Fang, Z., Yue, K., Zhang, J., Zhang, D., and Liu, W. (2014). Predicting click-through rates of new advertisements based on the Bayesian network. *Mathematical problems in engineering*.

[9] Galpin, J. and Gullen, P. (2000). Beyond the OTS: measuring the quality of media exposure. *International Journal of Market Research*, 42(4), 473-493.

[10] Jensen, F.V. and Nielsen, T. D. (2007). *Bayesian Networks and Decision Graphs*. New York: Springer, 2nd

ed.

[11] Katz, H. (2017). *The media handbook: A complete guide to advertising media selection, planning, research, and buying*. 6th ed. New York: Taylor & Francis.

[12] Kwoka, J. E. (1993). The sales and competitive effects of styling and advertising practices in the US auto industry. *The Review of Economics and Statistics*, 649-656.

[13] Lavidge, R. J. and Steiner, G. A. (1961). A model for predictive measurements of advertising effectiveness. *Journal of Marketing*, 25(6), 59-62.

[14] Novák, M. and Kincl, T. (2013). Television Commercials on Czech Market: Audience Targeting or Undifferentiated Approach? In: *Vision 2020: Innovation, Development Sustainability, and Economic Growth*. Wien, 27.06.2013 28.06.2013. USA: International Business Information Management, 1-9.

[15] Nelson, J. P. (2005). Beer advertising and marketing update: Structure, conduct, and social costs. *Review of Industrial Organization*, 26(3), 269-306.

[16] Pearl, J. and Russell, S. (2000). Bayesian Networks. *Department of Statistics, UCLA*. Available at: https://escholarship.org/uc/item/53n4f34m

[17] R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Available at: http://www.R-project.org/

[18] Rohloff, A. C. (1966). Quantitative analyses of the effectiveness of TV commercials. *Journal of Marketing Research*, 239-245.

[19] Rossi, P. E., Allenby, G. M. and McCulloch, R. (2005) *Bayesian Statistics and Marketing*. Chichester, West Sussex: John Wiley and Sons Ltd.

[20] Wind, Y. and Denny, J. (1974). Multivariate analysis of variance in research on the effectiveness of TV commercials. *Journal of Marketing Research*, 136-142.

[21] Yuspeh, S. (1977). *On-air: Are we testing the message or the medium?*. New York: J. Walter Thompson.

Food and drink	**Textile, clothing, leather, shoes**
news programme - main news session (75.0)	entertainment programme - show, stand-up (100.0)
entertainment programme - game, lottery (66.7)	news programme - main news session (100.0)
publicistic programme - talk show (44.4)	unspecified publicistic programme (71.4)
unspecified publicistic programme (39.5)	publicistic programme - talk show (50.0)
publicistic programme - free time and lifestyle (33.3)	publicistic programme - free time and lifestyle (33.3)
entertainment programme - show, stand-up (33.3)	second
Other manufacturing industry	**Finance and insurance**
news programme - special news (100.0)	entertainment programme - game, lottery (100.0)
publicistic programme - advices for consumers (80.0)	news programme - main news session (75.0)
news programme - main news session (75.0)	publicistic programme - topical general (50.0)
publicistic programme - journalistic magazine (75.0)	publicistic programme - talk show (43.0)
entertainment programme - game, lottery (50.0)	unspecified publicistic programme (41.2)
publicistic programme - topical general (50.0)	
Building industry	**Education**
unspecified sport programme (100.0)	unspecified entertainment programme (100.0)
Transport and storage	**Accommodation, food service and hospitality**
unspecified publicistic programme (100.0)	entertainment programme - thematic magazine (66.7)
entertainment programme - talk show (100.0)	
Information and communication activities	**Professional, scientific and technical services**
publicistic programme - advices for consumers (100.0)	dramatic programme - movie (50.0)
publicistic programme - journalistic magazine (50.0)	unspecified publicistic programme (50.0)
educational programme - programme for schools (50.0)	news programme - main news session (50.0)
entertainment programme - show, stand-up (50.0)	unspecified entertainment programme (33.3)
unspecified publicistic programme (44.4)	
Cultural, amusement and recreation activities	**Administrative and support services**
entertainment programme - reality show (100.0)	publicistic programme - journalistic magazine (50.0)
dramatic programme - TV series (33.3)	entertainment programme - competition (50.0)
dramatic programme - TV drama (33.3)	dramatic programme - sitcom (33.3)
Public administration and defence, compulsory social security	**Electricity, gas, heating and air conditioning**
entertainment programme - reality show (100.0)	publicistic programme - humanities (100.0)
Cultural, amusement and recreation activities	**Cross-sectional category**
publicistic programme - journalistic magazine (100.0)	news programme - special news (75.0)
publicistic programme - talk show (100.0)	news programme - main news session (66.7)
entertainment programme - talk show (100.0)	publicistic programme - journalistic magazine (50.0)
entertainment programme - show, stand-up (100.0)	publicistic programme - talk show (50.0)
unspecified publicistic programme (50.0)	entertainment programme - game, lottery (50.0)
entertainment programme - thematic magazine (50.0)	entertainment programme - quiz (50.0)
Other services	
news programme - main news session (100.0)	

Table 4 Probability of meaningful exposure given *Product category* and *Type of programme before a break* (in %)

A Contribution to the Application of the Fuzzy Approach to the Economic Analyses

Simona Hašková[1]

Abstract. The paper deals with the issue of a three-step multi-criterial evaluation of alternatives in the case of ex post evaluation. In the first stage the alternative A is assessed by the individual components of the specified vector of its basal data through which it is represented. In the second stage, the suitability of alternative A is assessed in the same way according to each of the values of the results of the partial single-criterion evaluations of alternative A. The third stage of evaluation is the number that is a sophisticated aggregate of the results of all single-criterion evaluations on a common scale, taking into account their mutual synergy (multi-criterial evaluation of the alternative A). Within this third stage, the uncertainty often comes into play arising from the vagueness of the description of the dependence of the resulting aggregate on the vector of the results values of all single-criterion evaluations. In a concrete example it is shown and demonstrated how a fuzzy approach based on fuzzy logic faces this internal uncertainty of the herein presented evaluation process of alternatives.

Keywords: fuzzy approach, multi-criterial evaluation, uncertainty, three step analysis.

JEL Classification: C51, C65
AMS Classification: 03E72, 90B90

1 Introduction

Methods for evaluating investments in production technologies play an important role in any competitive business environment. A careful preliminary analysis of the implications of an investment decision is a necessary basis on which it is possible to build a rational expectation of future development and on which the survival or non-survival of a production enterprise depends. This concerns, in particular, costly, long-term investments that include the purchase of modern manufacturing technologies such as group technologies, flexible production systems, computer-integrated production systems, etc., which contribute to improving production performance in terms of cost, productivity, reliability and quality. The issue of classification and methodology of standard approaches for assessing the economic benefits of investments is dealt with [3], [4], [5] from different points of view.

The traditional economic assessment of investment in production facilities was discussed, for example, in [9] and [10]. These methods are based on discounted cash flow, and in the case of valuation of investments in long-term assets, they are, in practice, the most frequently used one-criterion methods (in detail in [1]).

However, these standard approaches analysed in the above-mentioned works more or less ignore the specific role played by the human factor in investment activities [11]. They urge the manager not to deviate from the framework of a rooted tradition of standard thinking that identifies an understanding of a phenomenon with the ability to quantitatively analyse it as if it were a mechanistic system subjected to algebraic, differential or integral equations. The manager gets into a similar position as a computer programmer who has to adapt his mind to the „computer mind“ (to his operating system), which only understands numbers and can only handle numbers. In contrast, most management activities, including the investment activities, do not require exact figures, and, in many cases, such behaviour is even counterproductive by making it difficult for a manager to use his/her creative potential to its full extent [2].

[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.

For example, if we buy a car, we are interested in some quantitative parameters such as price, performance, consumption, etc., but we are also interested in some qualitative (aesthetic) parameters such as the colour and shape of the bodywork, which are linguistic variables characterized by natural language. We can choose well without knowing the exact wavelength of the colour shade and even without the knowledge of the three-dimensional system of equations of analytical geometry, precisely describing the shapes of the surfaces and the spatial structures from which the bodywork is compiled.

This is because the key elements of human thinking are not numbers, but intuitive concepts whose meanings and contents are vague, unclear, blurred, and simply fuzzy [6]. The human mind is gifted with the almost miraculous ability to use this fuzzy-intuitive concept to reduce a lot of data attacking the human brain to the minimally precise handful that is needed to carry out the task. This human ability is well understood by the authors working with the word fuzzy in the title of their papers given in the list of reference of this paper that partly builds on them. They did not choose the way, which forces the manager to „think like a computer", but vice versa; they attempt to adapt the methods of economic analysis to the manager´s natural style of thinking [8]. Then, unsurprisingly, the formal model of an intuitive concept will be a fuzzy set.

Given the existence of the qualitative effects of many investment activities, fuzzy linguistic decision-making will be discussed and applied as an effective alternative to the techniques, which ignore the role of human factor. The methodological part will be devoted to the basics of the fuzzy approach, based on which an original model of three-step multi-criteria assessment is constructed. In the application section this model will be applied to solve a similar problem, which we face when investing in the purchase of our own car - to evaluate and optimize the choice of production equipment, whose effects are of a financial and qualitative nature.

2 The methodological approach

The *fuzzy approach* has its bases in different versions of fuzzy logic that have been created by adapting the binary numerical characteristics of propositional operators to the interval $\langle 0,1\rangle$. If $|A|\in\{0,1\}$, or respectively $|B|\in\{0,1\}$ being the truth value of the statement A, or respectively B, we can classify the classical truth tables for negation, conjunction, disjunction, and implication by the numerical characteristics of these operators: $|\neg A| = 1 - |A|$, $|A\wedge B| = \min\{|A|,|B|\}$, $|A\vee B| = \max\{|A|,|B|\}$ and $|A\rightarrow B| = 1$ for $|A| \leq |B|$, $|A\rightarrow B| = 1 - (|A| - |B|)$, otherwise. Moving from the set $\{0,1\}$ to the interval $\langle 0,1\rangle$ we get the semantic form of Lukasiewicz's fuzzy propositional calculus originally formulated by Lukasiewicz and Tarski [12]. For us, the fuzzy logic is only a tool for exact dealing with fuzzy sets, whose theory was published by L. A. Zadeh – see [13] and [14].

2.1 Elements of fuzzy set theory

Let the set U be a field of consideration or discussion (universe), let $\mu_{\underline{A}}$: $U \rightarrow \langle 0,1\rangle$ be a membership function and let $\underline{A} = \{(y, \mu_{\underline{A}}(y)): y \in U\}$ be the set of all pairs $(y, \mu_{\underline{A}}(y))$ in which the numbers $0 \leq \mu_{\underline{A}}(y) \leq 1$ indicate the degree of affiliation of the pair $(y, \mu_A(y))$ to the set $\underline{A}$ to the given $y \in U$. Then $\underline{A}$ is a *fuzzy subset* on the universe U. The significant characteristic of the fuzzy subset $\underline{A}$ is its support $U_{\underline{A}} = \{y: 0 < \mu_{\underline{A}}(y) \leq 1, y \in U\} \subset U$. In terms of fuzzy logic is $\mu_{\underline{A}}(y) = |y \in U_{\underline{A}}|$. The element $y \in U$ with $\mu_{\underline{A}}(y) = 0.5$ is called the *crossover point* in $\underline{A}$. At the values greater than 0.5 the element y *rather* belongs to $U_{\underline{A}}$, at the values smaller than 0.5 it *rather* does not belong to $U_{\underline{A}}$.

Fuzzy subset $\underline{A}$, whose carrier $U_{\underline{A}} \subset U \subset R$, where R is the set of the real numbers, and its function $\mu_{\underline{A}}$ is gifted by the property of normality and convexity, i.e. at least in the case of one element $x \in U_{\underline{A}}$ it applies for it $\mu_{\underline{A}}(x) = 1$, and $\mu_{\underline{A}}(x') \geq \min\{\mu_{\underline{A}}(x_1), \mu_{\underline{A}}(x_2)\}$for all $x' \in \langle x_1, x_2\rangle \subset U_{\underline{A}}$, is called the *fuzzy number*. Theoretically, different forms of the membership function $\mu_{\underline{A}}$ of fuzzy numbers can exist: triangular, trapezoidal, bell-shaped, sinusoidal, cosinusoidal (see e.g. [7] and [15]).

In this paper, fuzzy numbers are formal representatives of values (terms) of linguistic variables. A typical linguistic variable is, for example, *light*. From the physical point of view, it is for the human eye a visible part of electromagnetic radiation with wavelengths between 380 and 750 nm (the field of basal numerical values of the linguistic variable – universe U). The values (terms) of the linguistic variable *light*

are the basic colours of the rainbow, i.e. *yellow*, *orange*, *red*, *green*, *blue*, *indigo* and *purple* (see [16]). In points where the primary colours blend, their colour shades (e.g. magenta, pink, etc.) are created. This matter is best reflected by a model of seven trapezoid fuzzy numbers (two border and five inner numbers) whose carriers cover the universe U and at their ends they intersect with adjacent carriers.

For the needs of managerial decision-making a rougher model usually suits with one inner and two border fuzzy numbers for the terms *low* (*L*), *common* (*M*) and *high* (*H*) value. Interval U is divided by the points a, b, c, d into five sections with the membership functions (1), (2) and (3):

$$\mu_{\underline{L}}(x) = 1 \; for \; x < a, \mu_{\underline{L}}(x) = \frac{b-x}{b-a} \; for \; a \leq x < b, \mu_{\underline{L}}(x) = 0 \; otherwise. \tag{1}$$

$$\mu_{\underline{M}}(x) = \frac{x-a}{b-a} \; for \; a \leq x < b, \mu_{\underline{M}}(x) = 1 \; for \; b \leq x < c, \mu_{\underline{M}}(x) = \frac{d-x}{d-c} \; for \; c \leq x < d, \mu_{\underline{M}}(x) = 0 \; otherwise. \tag{2}$$

$$\mu_{\underline{H}}(x) = 0 \; for \; x < c, \mu_{\underline{H}}(x) = \frac{x-c}{d-c} for \; c \leq x < d, \mu_{\underline{H}}(x) = 1 \; otherwise. \tag{3}$$

The position of the points a to d in the universe U is determined by an expert in the field of knowledge.

Another important tool of fuzzy set theory is the *rule*. It occurs in two modes: in the *symbolic* notation and in the *fuzzy logical* notation.

In the *symbolic* notation, there is the functional relation between the n-member vector of values (fuzzy numbers $\underline{A}_i$ on universes U_i, i = 1,…, n) of different linguistic variables standing on the left side of the rule and the value of linguistic variable (fuzzy number $\underline{B}$ on universe V), standing on the right side of the rule, expressed as a pair $((\underline{A}_{1,\dots,}\underline{A}_n), \underline{B})$. The set of rules $P = \{((\underline{A}_{1,\dots,}\underline{A}_n), \underline{B})$: $\underline{A}_i \in \{\underline{L}_i, \underline{M}_i, \underline{H}_i\}$, $\underline{B} \in \{\underline{L}, \underline{M}, \underline{H}\}$, i = 1,…, n$\}$, whose right sides were chosen by an expert, is a set of *inference* rules and works just like a discrete function of n variables in mathematical analysis (it transforms the n-dimensional vectors of fuzzy numbers standing on the left sides of the rules into fuzzy numbers standing on the right sides of the rules).

The *fuzzy logical* notation takes into account the adequacy of the rule in the situation characterized by the vector $x = (x_1,\dots, x_n)$ by interpreting the vector $(\underline{A}_{1,\dots,}\underline{A}_n)$ of the left side of the rule as a fuzzy logical conjunct of statements $(x_1 \in U_{\underline{A}1}) \wedge \dots \wedge (x_n \in U_{\underline{A}n})$, whose truth value $|(x_1 \in U_{\underline{A}1}) \wedge \dots \wedge (x_n \in U_{\underline{A}n})| = \min\{|(x_1 \in U_{\underline{A}1})|, \dots, |(x_n \in U_{\underline{A}n})|\} = \min\{\mu_i(x_i), \dots, \mu_i(x_i)\}$ is the measure of its adequacy. This calculated value limits the course of the function $\mu_{\underline{B}}$ on the universe V from the above of the fuzzy number standing on the right side of the rule (for details see the following part).

2.2 Construction of a three-step fuzzy model of multi-criteria evaluation

The three-step fuzzy multi-criteria model describes a multi-criteria evaluation of final set of alternatives. Let A be a given set of considered alternatives, specified by vectors a of their basal data. Then, as the multi-criteria evaluation of the alternative specified by the vector $a = (a_1,\dots,a_c) \in A$ we consider the number y_x, which is the output of the process, whose three-step general model is schematically shown in the left part of Fig. 1.

Before the multi-criteria evaluation itself, we assess the alternative a at the first stage of the process according to the values of its basal data in comparison to other alternatives. If the preference order is not decided at this stage, the vector a will be subjected to a number of different single-criterion evaluations at the second stage. If the comparative method does not decide on the order of advantage at that stage, then evaluation at the third stage (block K) will follow. In the right part of Fig. 1, there is a case in which some basal data enter the block K directly.

The fuzzy approach applied in the third stage of the process consists of four phases:

- *Fuzzification* – the input vector $x = (x_1,\dots, x_n)$ converts each inference rule from the set P into the *logical* notation mode, i.e. in the form $(\min\{\mu_1(x_1),\dots, \mu_n(x_n)\}, \mu_{\underline{B}})$, where $\mu_i(x_i) = |x_i \in U_{\underline{A}i}|$; the left side of the rule in the logical notation is the number, the right side is the function $\mu_{\underline{B}}$: $V \rightarrow \langle 0,1 \rangle$. The set P is thus transformed to $P^* = \{(\min\{\mu_1(x_1),\dots, \mu_n(x_n)\}, \mu_{\underline{B}})$: $\min\{\mu_1(x_1),\dots, \mu_n(x_n)\} > 0\}$.

- *A set of partial results* is a set $B = \{\min\{\min\{\mu_1(x_1),\ldots, \mu_n(x_n)\}, \mu_{\underline{B}}\}: (\min\{\mu_1(x_1),\ldots, \mu_n(x_n)\}, \mu_{\underline{B}}) \in P^*\}$. It is a mixture of $\mu_{\underline{B}}$ functions whose curves over the universe V are limited from the above by the numbers $\min\{\mu_1(x_1),\ldots, \mu_n(x_n)\}$, taking into account the degree of adequacy of the rule application. These partial results are projected into a common scale (universe V), and it is necessary to incorporate them into the common output fuzzy set by a logical sum (aggregation) and to determine its affiliation function. This will happen in the next phase of the solution process.
- *Aggregation* or summary of functions of set B into a compact unit and detection of its μ_{agg}. Using the so-called extension principle (see e.g. [12]), which is one of the basic concepts of the fuzzy set theory enabling us to find out that this compact set is a fuzzy subset on the universe V with $\mu_{agg} = \max\{\min\{\min\{\mu_1(x_1),\ldots, \mu_n(x_n)\}, \mu_{\underline{B}}\}: (\min\{\mu_1(x_1),\ldots, \mu_n(x_n)\}, \mu_{\underline{B}}) \in P^*\}$.
- *Defuzzification* is the final phase of the process, in the last step of which we obtain the output value y_x of the three-step model of the multi-criterial evaluation of the alternative $(a_1,\ldots,a_c) \in A$ as the mean value of the elements $y \in V$ weighted by the values $\mu_{agg}(y)$ of its importance, and therefore (4) it applies:

$$y_x = \int y \cdot \mu_{agg}(y)dy / \int \mu_{agg}(y)dy, \tag{4}$$

where $\int$ is the symbol of a certain integral over the universe Y.

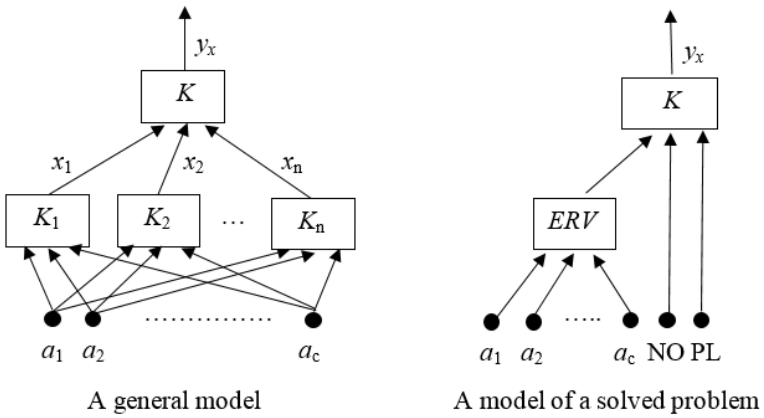


Figure 1 Structure of a three-step model of multi-criteria evaluation of alternatives, where NO and PL stands for specific components of a vector a (see part 3.1)

3 Application of a three-step fuzzy model to the problem of production facility choice

The key factors of the specific cases that need to be taken into account when introducing new production technology can be summarized as follows: selection of the type of production facility, selection of the time of its decommissioning, the way of its subsequent usage, purchase price, cost and yield characteristics, qualitative effects. In addition, the relevant qualitative criteria are often taken into account.

3.1 Task assignment

The manager has to decide about a choice between two suitable production facilities (machines) with lifetimes of 6 years and 9 years. For each machine the relevant numerical data recorded in the managerial accounting are provided, from which can be found: the acquisition cost and annual operating expenses assuming standard use throughout the physical lifetime including the tax rate on the profit (see Table 1) and cost of capital of the project (discount rate 10 %) in which the production facility is involved. The data on the annual income from the use of the production facility are the same in both cases and therefore irrelevant in terms of machine choice.

Year	0	1	2	3	4	5	6	7	8	9
Technology X	31.0	7.9	13.4	18.9	26.4	36.3	52.8	-	-	-
Technology Y	44.3	5.6	11.0	13.3	16.5	19.8	24.42	30.4	39.6	52.8

Table 1 Capital and annual expenditures for technology X (n = 6 years) and Y (n = 9 years) in thousands of EUR

The two parameters of a qualitative nature that have to be taken into account are the noise of manufacturing operation (NO) of the production facility and its impact on the pollution level (PL) of the surrounding environment. According to available knowledge, both parameters are more favourable with technology X: the NO´s level lies between the low and normal noise level, with the technology Y is the NO is rather normal than high. PL caused by the operation of the technology X is rather normal than high, with the machine Y, PL lies on the border between the normal and the high level.

The input parameters of the alternatives are of minimizing nature from the viewpoint of a decision-maker; in the next, they are processed within the three-step fuzzy model. The first stage stands for a simple comparison of data alternatives (here: investment, annual operating expenses, NO, PL). In case of impossibility to decide on the alternative, the solution proceeds to the second stage in which the set of inputs is modified by means of a relevant result / results of an one-criterion evaluation; these inputs are compared again within each alternative (here, the inputs at the second stage: the equivalent of annual expenditures, NO, PL). If optimal choice is not possible, the solution proceeds to the third stage, in which the fuzzy approach described by the four-phase process in Section 2.2 is applied.

3.2 Task solution

From a simple comparison of input data, the problem cannot be decided *at the first stage* of our three-step model. The technology X requires a lower initial investment, has more favourable NO and PL characteristics and mostly higher annual operating costs; the operating costs are an irrelevant decision source, since machine X differs from machine Y in lifetime. For that reason the average annual expenditures (including initial investment) are calculated as the equivalent of annual expenditures ERV = PV / α, where PV is the current value of operating expenditures stated in Table 1, and α is the relevant annuity factor. If $ERV_X \leq ERV_Y$, the problem would be resolved *at the second stage* of our model with the technology X being the optimal choice. Unfortunately, this is not the case: ERV_X = 30.7 thousand EUR, ERV_Y = 28.1 thousand EUR. The decision will be made *at the third stage* of the model (see the diagram in the right part of Figure 1).

For this purpose, we normalize the values of inputs to block *K* by specifying them in the percentage range of their universe. If the universe *U* of values of the variable ERV is between 0 and 50 thousand EUR, then $ERV_X \approx 62$ % and $ERV_Y \approx 56$ %. Estimation of the parameter values NO and PL depends on the distribution of points a to d in the definition of the trapezoid fuzzy numbers (see part 2.1 relations (1), (2), (3)) and from it resulting crossover points. If there is no expert who could help us with the distribution of points a to d, it is best to assume that they are distributed evenly, i.e. a = 20 %, b = 40 %, c = 60 % and d = 80 %. Then NO_X = 30 %, $NO_Y \approx 65$ %; $PL_X \approx 65$ %, PL_Y = 70 %. The set of inferential rules $P = \{((\underline{ERV}, \underline{NO}, \underline{PL}), \underline{B})$: $\underline{ERV}, \underline{NO}, \underline{PL}, \underline{B} \in \{\underline{L}, \underline{M}, \underline{H}\}\}$ has 27 elements in total, in which the selection of the right sides (without help of an expert) is performed by the „predominant element“ method in the left side vector (if neither of the two of them coincide, then $\underline{B} = \underline{M}$).

For technology X we get $P^* = \{(0.5, \mu_{\underline{M}}), (0.25, \mu_{\underline{M}}), (0.1, \mu_{\underline{M}}), (0.1, \mu_{\underline{H}})\}$, $B = \{\min\{0.5, \mu_{\underline{M}}\}, \min\{0.25, \mu_{\underline{M}}\}, \min\{0.1, \mu_{\underline{M}}\}, \min\{0.1, \mu_{\underline{H}}\}\}$, $\mu_{agg} = \max\{\min\{0.5, \mu_{\underline{M}}\}, \min\{0.1, \mu_{\underline{H}}\}\}$, therefore $\mu_{agg}(y) = 0$ for $y < 20$, $\mu_{agg}(y) = (y - 20) / 20$ for $20 \leq y < 30$, $\mu_{agg}(y) = 0.5$ for $30 \leq y < 70$, $\mu_{agg}(y) = (80 - y) / 20$ for $70 \leq y < 78$, $\mu_{agg}(y) = 0.1$ for $y \geq 78$. By integration by parts (see (4)) we get $y_x = \int y \cdot \mu_{agg}(y)dy / \int \mu_{agg}(y)dy = 1437.9 / 27.1 = 53.06$.

For technology Y we get $P^* = \{(0.5, \mu_{\underline{M}}), (0.25, \mu_{\underline{M}}), (0.25, \mu_{\underline{H}})\}$, $B = \{\min\{0.5, \mu_{\underline{M}}\}, \min\{0.25, \mu_{\underline{M}}\}, \min\{0.25, \mu_{\underline{H}}\}\}$, $\mu_{agg} = \max\{\min\{0.5, \mu_{\underline{M}}\}, \min\{0.25, \mu_{\underline{H}}\}\}$, therefore $\mu_{agg}(y) = 0$ for $y < 20$, $\mu_{agg}(y) = (y - 20) / 20$ for $20 \leq y < 30$, $\mu_{agg}(y) = 0{,}5$ for $30 \leq y < 70$, $\mu_{agg}(y) = (80 - y) / 20$ for $70 \leq y < 75$, $\mu_{agg}(y) = 0.25$ for $y \geq 75$. By integration by parts (see (4)) we get $y_x = \int y \cdot \mu_{agg}(y)dy / \int \mu_{agg}(y)dy = 1749.1 / 30.625 = 57.1$.

Based on the results of performed fuzzy analysis it is concluded that technology X is slightly more advantageous than its alternative Y as the mean value in terms of (4), covering average annual expenditures and "harm" resulting from the usage of the technology expressed in thousands of EUR, is smaller.

4 Summary and conclusion

The introduction part stated the reasons that support the tendency to use fuzzy approach in managerial analyses. The methodological part covers a brief insight into the basics of philosophy, theory and techniques of fuzzy approach that is necessary for understanding the author´s concept of fuzzy multi-criteria evaluation presented in the form of a three-step model. In the final part this model was applied to the problem of the optimal choice of a particular production facility. Given the input data and available information of qualitative nature, the result of the fuzzy analysis indicated that technology X is slightly more advantageous than its alternative Y.

References

[1] Brealey, R. A., Myers, S. C. & Marcus, A. J. (2011). *Fundamentals of Corporate Finance*. Mcgraw-Hill Education, New York.

[2] Corbin, J. C., Reyna, V. F., Weldon, R. B. & Brainerd, C. J. (2015). How reasoning, judgment, and decision making are colored by gist-based intuition: A fuzzy-trace theory approach. *Journal of applied research in memory and cognition*, ***4***(4), 344-355.

[3] Fiala, P. & Borovička, A. (2015). Investment decision-making by a two-step multi-criteria procedure. *International Journal of Multicriteria Decision Making*, ***5***(3), 274-293.

[4] Hašková, S. (2016). Evaluation of Project Investments Based on Comprehensive Risk. In A. Kocourek & M. Vavroušek (Eds.), *Proceedings of 34th International Conference Mathematical Methods in Economics (MME)* (*pp. 260-264*). Liberec: Technical University Liberec.

[5] Hasek, R., & Zahradnik, J. (2015). Study of historical geometric problems by means of CAS and DGS. *The International Journal for Technology in Mathematics Education*, ***22***(2), 53-59.

[6] Herrera, F. & Herrera-Viedma, E. (2000). Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and systems*, ***115***(1), 67-82.

[7] Kahraman, C. (2001). Capital budgeting techniques using discounted fuzzy cash flows. In D. Ruan, J. Kacprzyk & M. Fedrizzi (Eds.), *Soft Computing for Risk Evaluation and Management. Studies in Fuzziness and Soft Computing*, ***76***, 375-396. Physica, Heidelberg.

[8] Karsak, E. E., & Tolga, E. (2001). Fuzzy multi-criteria decision-making procedure for evaluating advanced manufacturing system investments. *International journal of production economics*, ***69***(1), 49-64.

[9] Meredith, J.R. & Suresh, N.C. (1986). Justification techniques for advanced manufacturing technologies. *International Journal of Production Research*, **24**(5), 1043-1057.

[10] Proctor, M.D. & Canada, J.R. (1992). Past and present methods of manufacturing investment evaluation: A review of the empirical and theoretical literature. *The Engineering Economist*, **38**(1), 45-58.

[11] Ross, T. J. (2010). *Fuzzy logic with engineering applications.* Chichester: John Wiley & Sons, Ltd, 3rd Ed.

[12] Svoboda, V. & Peregrin, J. (2009) *Od jazyka k logice: Filozofický úvod do modern logiky*. Academia, Praha.

[13] Zadeh, L. A. (1975). Fuzzy logic and approximate reasoning. *Synthese*, ***30***(3-4), 407-428.

[14] Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, **8**(3), 338-353.

[15] Zhao, J. & Bose, B. K. (2002). Evaluation of membership functions for fuzzy logic controlled induction motor drive. In *IECON-2002: Proceedings of the 2002 28 the Annual Conference of the IEEE Industrial Electronics Society* (*Vol. 1: pp. 229-234*). New York: IEEE Industrial Electronics Society.

[16] Wikipedia, The Free Encyclopedia (2018). Light. [2018-10-4]. Available from: https://en.wikipedia.org/wiki/Light

Common shock approach to default risk of reinsurance: Solvency II framework

Radek Hendrych[1], Tomáš Cipra[2]

Abstract. This contribution deals with the construction of required capital covering the default risk in portfolios with a small number of heterogeneous counterparties. The typical application of this concept is the counterparty default risk of reinsurance in the Solvency II framework. A special approach to default risk based on the common shock principle is investigated in greater detail. In particular, a numerical study compares results of various methods which are applicable in this context. The numerical results demonstrate that the suggested modifications of the widely accepted common shock approach implemented within the official Solvency II regulation might be preferred by insurance companies when constructing the portfolio of reinsurers for reinsurance protection of given insurance company. Moreover, it is also confirmed that a single high-quality counterparty is more effective in the context of counterparty default risk than a number of low-quality counterparties.

Keywords: default risk, common shock, Solvency II, value-at-risk.

JEL Classification: G22
AMS Classification: 62P05

1 Introduction

The paper deals with the counterparty default risk for a given insurer, which is mainly the risk of possible defaults in the portfolio of reinsurers of this insurer. The counterparty default risk plays an important role in the new regulatory system Solvency II introduced for the European insurance industry since 2016. The regime Solvency II covers "all risks" and not only traditional risk categories like underwriting risk or market risk (see e.g. [1] or [8]). Particular risks are grouped to risk modules (sorted further to distinct risk submodules) with corresponding solvency capital requirements (SCR) calculated either by standard formulas or by internal models approved by insurance supervisors. The Solvency II prescribes that the insurer must cover all risk by own eligible funds calculated just according to SCR.

The paper focuses on modeling approaches to the counterparty default risk of reinsurance. Since the number of reinsurers in the reinsurance portfolio of one insurer is limited (rarely more than ten reinsurers) one cannot replicate typical credit risk modeling approaches which are typical for banks. Such modeling concepts are based on the key assumption of large homogenous population of debtors which enables to apply simplifying asymptotic results (e.g. the broadly applied Vasicek limiting probability distribution, refer to [10]).

There are other two aspects that should be reminded in the situation of reinsurance portfolio with limited number of counterparties: The first of them consists in the fact that the values of loss given default (LGD) of particular reinsurers must be estimated in a sophisticated way (undoubtedly these are very important inputs for the calculation of SCR to cover counterparty default risk of reinsurance). Since this aspect is a very empirical concern and includes conclusions and recommendations from corresponding quantitative impact studies we avoid it in this text (see e.g. [4] or [5]). The second aspect is related to the problem of diversification and correlation in the studied reinsurance portfolios. Modeling correlation of defaults plays a prominent role in pricing and risk management generally.

A promising approach to modeling correlation in reinsurance portfolios is based on a common shock. It generates the correlation among reinsurers in a specific way (see [7] or [9]). Similarly as the insurance industry, the reinsurance industry is vulnerable to critical phenomena of sudden impacts induced by natural catastrophes, financial crises, terrorism, etc. Using this principle, the correlation structure follows from structural parameters of a special model that is constructed for the common shock impact in the given reinsurance portfolio. Since this approach respects the heterogeneity of reinsurers with different default probabilities and can be applied for small number of reinsurers in the portfolio, one recommends it in this context for calculating the particular SCR.

[1] Charles University, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague, Czech Republic, hendrych@karlin.mff.cuni.cz.
[2] Charles University, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague, Czech Republic, cipra@karlin.mff.cuni.cz.

The aim of this paper is to survey the Solvency II approach to counterparty default risk of reinsurance based on the common shock principle, suggest other possible methods and compare them by means of a numerical study (we usually assume the simple model situation with the uniform distribution of loss given defaults in the portfolio to be compatible with other authors). In particular, we deal with the problem how to calculate the *required capital* (*RC*) covering the counterparty default risk due to *n* risk expositions. It must be retained by an insurance company in accordance with Solvency II. It is worth noting that we use the more general term *required capital* instead of *solvency capital requirements SCR* to stress the fact that more general applications are possible: Although the paper deals primarily with measuring counterparty default risk in the context of reinsurance, the results can be applied also in other situations with a small number of heterogeneous risk expositions.

The paper is organized as follows. Section 2 describes the common shock approach to counterparty default risk of reinsurance. Section 3 studies methods for prescribing the required capital by employing the common shock concept. Section 4 presents the results of numerical study. Section 5 contains conclusion.

2 Common shock approach to counterparty default risk of reinsurance

The common shock approach in reinsurance was introduced in [9]. It can be described as follows:

Let us denote the common shock affecting *n* reinsurers of a given insurance company (e.g. the arrival of the financial and economic crisis or recession, a legislative change or reform, a catastrophic event) as a random variable *R* ranging in the interval between zero and one. For the values of *R* near to zero or one, the common influence of the given phenomenon is low or high, respectively. The behavior of *R* can be described by the probability density function of the form:

$$f(r) = \beta r^{\beta-1}, \quad 0 < r < 1, \tag{1}$$

where β is a parameter ($0 < \beta < 1$). Such a probability density function can be looked upon as a special case of the beta distribution, and it is acceptable from the practical point of view since according to (1) the small shocks are the most probable ones, while the probability of more intensive shocks declines to small positive values (however, the zero value is never achieved). Another interpretation is also possible: the maximal shock among *n* mutually independent annual shocks (i.e., during *n* years) has the probability density function of the form (1) but with the parameter $n\beta$ instead of β. Therefore e.g. for $\beta = 0.1$ and $n = 10$, the probability density function (1) corresponds to the uniform distribution so that the maximal shock during the period of 10 years attains each of its values with the same probability (one can make use of this property when calibrating β).

The probabilities of default will depend on the common shock *R*. A suitable functional relation seems to be:

$$PD_i(r) = p_i + (1 - p_i) r^{\gamma / p_i}, \quad 0 < r < 1, \tag{2}$$

where p_i is a basic level of the probability of default of the *i*-th counterparty ($i = 1, \ldots, n$) in the reinsurance portfolio (it is a benchmark after excluding the influence of the common shock *R*). Moreover, in (2) one adds to this basic level a component that depends on *R* by means of a positive parameter γ. The power exponent in (2) is a decreasing function of the basic level of the probability of default p_i since counterparties with a low p_i are not very sensitive to the random shocks while the higher probabilities of default increase the default sensitivity even if the attained values of *R* are not high. In any case, the function (2) increases from the basic level p_i to the value 1. Moreover, this function is concave for $\gamma < p_i$ (in such a case, the probability of default PD_i is regarded as large), it is convex for $\gamma > p_i$ (then the probability PD_i is regarded as small); for $\gamma = p_i$ one obtains an increasing line.

By integrating the function (2) over *r* using the probability density function (1) one obtains the following formula for the probability of default of the *i*-th counterparty as:

$$PD_i = \mathrm{E}(PD_i(R)) = \int_0^1 PD_i(r) f(r) dr = \frac{(\gamma + \beta) p_i}{\gamma + \beta p_i} = \frac{(\beta/\gamma + 1) p_i}{1 + \beta/\gamma \cdot p_i}. \tag{3}$$

Conversely, if one takes the numerical values of default probabilities PD_i of particular counterparties as an external rating provided by specialized agencies (see e.g. Standard & Poor's ratings in Table 1) then it is possible to express the basic level of the probability of default p_i of the *i*-th counterparty evidently as:

$$p_i = \frac{\gamma\, PD_i}{\beta(1 - PD_i) + \gamma} = \frac{\gamma/\beta \cdot PD_i}{(1 - PD_i) + \gamma/\beta}. \tag{4}$$

One can summarize that the behavior of particular counterparties in the credit portfolio may be modeled in a suitable parametric way using observed values PD_i and the ratio of two parameters β and γ. The higher the ratio β/γ is, the more does the observed probability of default differ from the baseline probability of default. For instance, from (3.3) it follows that $PD_i \approx 1.5 \cdot p_i$ for $\beta/\gamma = 0.5$ and $PD_i \approx 5 \cdot p_i$ for $\beta/\gamma = 4$ (if p_i is close to zero).

Finally, one extends the results for particular counterparties to the whole credit portfolio. Let I_i is a random default indicator of the i-th counterparty, i.e., $I_i = 1$ or $I_i = 0$ depending whether the default has occurred or not, respectively. Then the loss L generated by the defaults in the whole portfolio can be expressed in the form:

$$L = \sum_{i=1}^{n} LGD_i \cdot I_i \tag{5}$$

(LGD_i is the particular loss following from the default of the i-th counterparty). It holds (refer to [2]):

$$\mathrm{E}(L) = \sum_{i=1}^{n} LGD_i \cdot PD_i, \quad \mathrm{var}(L) = \sum_{i=1}^{n} \sum_{j=1}^{n} LGD_i \cdot LGD_j \cdot \sigma_{ij}, \tag{6}$$

where

$$\sigma_{ii} = PD_i(1 - PD_i), \quad \sigma_{ij} = \frac{\beta(1-p_i)\cdot(1-p_j)}{\beta + \gamma p_i^{-1} + \gamma p_j^{-1}} - (PD_i - p_i)\cdot(PD_j - p_j), \quad i \neq j \cdot \tag{7}$$

AAA	AA	A	BBB	BB	B	CCC+
0.002%	0.010%	0.050%	0.240%	1.200%	6.040%	30.410%

Table 1 Standard & Poor's (S&P) ratings and corresponding probabilities of default (PD)

3 Calculating required capital

3.1 Value-at-risk based on normal distribution in common shock model

Ter Berg [9] in his seminal work suggested to calculate the required capital RC by means of the common shock approach in the simplest possible way, i.e. as the (relative mean corrected) value-at-risk VaR based on the approximation by normal distribution using the corresponding moments (6) – (7) as:

$$RC_\alpha = \min\left\{\sum_{i=1}^{n} LGD_i, \Phi^{-1}(\alpha)\cdot\sqrt{\mathrm{var}(L)}\right\} \tag{8}$$

for $0 < \alpha < 1$. $\Phi^{-1}(\alpha)$ denotes the $(\alpha \times 100)$%-quantile of the standard normal distribution.

This suggestion was later implemented in QIS5 (2011) in the adjusted form:

$$RC_\alpha = \min\left\{\sum_{i=1}^{n} LGD_i, q\cdot\sqrt{\mathrm{var}(L)}\right\} \quad \text{with} \quad q = \begin{cases} 3, & \sqrt{\mathrm{var}(L)} \le 0.05\cdot\sum_{i=1}^{n} LGD_i, \\ 5, & \sqrt{\mathrm{var}(L)} > 0.05\cdot\sum_{i=1}^{n} LGD_i. \end{cases} \tag{9}$$

3.2 Value-at-risk based on true distribution in common shock model

The random variable L, i.e. the total loss due to counterparty portfolio defaults, has a discrete distribution (possibly with a very small number of discrete values). Therefore, its approximation by the normal approximation in (8) is very artificial. The opposite extreme is to calculate the corresponding value-at-risk exactly as the quantile:

$$RC_\alpha = F_L^{-1}(\alpha) \tag{10}$$

for $0 < \alpha < 1$, where $F_L(x)$ is the cumulative distribution function of L. In general, the random variable L acquires 2^n values $\{0, LGD_1, \ldots, LGD_n, LGD_1 + LGD_2, \ldots, LGD_1 + LGD_n, \ldots, LGD_1 + \ldots + LGD_n\}$, where e.g.

$$\mathrm{P}(L=0)=\int_0^1 \mathrm{P}(L=0\,|\,R=r)\,f(r)\,\mathrm{d}r=\int_0^1\left(\prod_{i=1}^{n}[1-PD_i(r)]\right)f(r)\,\mathrm{d}r=\int_0^1\left(\prod_{i=1}^{n}[1-p_i-(1-p_i)r^{\gamma/p_i}]\right)\beta r^{\beta-1}\,\mathrm{d}r \qquad (11)$$

according to (1) and (2) making use of the fact that the summands in (5) are conditionally independent for fixed values of the common shock intensity $R = r$. One can proceed in a similar way to derive the corresponding probabilities for other values of L but in general, their analytical expression is too complex.

Another possible approach is to calculate the corresponding quantile of L by means of simulations for the empirical distribution function of L, $\hat{F}_L$, i.e.:

$$RC_\alpha=\hat{F}_L^{-1}(\alpha)\,. \qquad (12)$$

Conditionally on the common shock size $R=r$ the true probability distribution of random loss L given by (5) follows from convolution of independent risks $\{I_i(r)\}_{i=1,\ldots,N}$, which are modeled by the alternative distribution with the corresponding probabilities of default $\{PD_i(r)\}_{i=1,\ldots,N}$. Weighing these conditional discrete distributions by the probability of common shock intensity delivers the true probability distribution of the random variable L.

To be more specific, the approximation of distribution function of L is calculated as follows: For given N, $\{LGD_i\}_{i=1,\ldots,N}$ and $R=r$ (which is repeatedly drawn from the distribution function of R), one can simulate the conditional distribution of L given $R=r$ (in accordance with the sum (5)), where $\{I_i(r)\}_i$ are independently drawn from the alternative distribution with the default probabilities $\{PD_i(r)\}_i$. Weighing these simulated conditional discrete distributions with the discretized estimated probability of the common shock R one obtains the approximate simulated probability distribution of the random loss L.

3.3 Expected shortfall in common shock model

The expected shortfall (ES) may be also applied in the framework of common shock approach. If the true distribution should be used similarly as in Section 3.2 then, it holds:

$$RC_\alpha=\frac{1}{1-\alpha}\int_\alpha^1 F_L^{-1}(u)\,\mathrm{d}u \quad \text{or} \quad RC_\alpha=\frac{1}{1-\alpha}\int_\alpha^1 \hat{F}_L^{-1}(u)\,\mathrm{d}u \qquad (13)$$

for $0 < \alpha < 1$, where one makes use of (10) and (12), respectively. Both integrals in (13) are computed numerically.

4 Comparative numerical study

In this section, we shall compare the methods for determining the required capital covering the counterparty default risk discussed in the previous sections. We shall analyze the portfolio of N counterparties (all with the equal probability of default) to be consistent with other published works (see e.g. [3]).

Figure 1 displays $RC_{0.995}$ for the portfolio of N similar risk expositions with equal credit rates and $LGD_i = 1/N$ for $i = 1,\ldots,N$ ($N = 1,\ldots,20$) calculated using the formulas in Section 3. Note that we put $\beta/\gamma = 4$, as it is recommended in [4], and used the probabilities of default given in Table 1. At first sight, the differences between results obtained by the true value-at-risk method (10) and its simulated counterpart (12) or between the true expected shortfall model and its simulated counterpart given by (13), respectively, are negligible. It means that the simulation strategy seems to be relatively precise. Point out that we estimated the conditional distribution $L\,|\,R=r$ by 1000 simulations for each given r and the discretized probability of common shock $R = r$ by 2000 simulations (in all relevant tasks). Refer to Section 3.2 for more details. Moreover, the value-at-risk based approaches in (8) and (9) tend to be generally inconsistent with the true and simulated value-at-risk method (10) and (12), respectively (even if one adequately adjusts the results for (8) and (9) by the term E(L) instead of the mean corrected VaR).

Firstly, it seems that the assumption of using the quantile of standard normal distribution in (8) is quite artificial (at least for smaller portfolios). It is confirmed by the detailed analysis of the overall shape of probability distribution of L: the random variable L has a firm probability mass at zero followed by an uneven tail to the right. Its skewness and kurtosis are substantially different from the standard normal ones, which significantly influences the speed and quality of convergence guaranteed by the central limit theorem. These facts lead to the (relatively high) inaccuracy of the approximation by the standard normal distribution. Therefore, the value-at-risk model (8) is not very adequate (at least for smaller portfolios). Secondly, analogous statements are valid also for the value-at-risk method (9), which uses the quantile factors 3 or 5 instead of the 99.5% quantile of standard normal distribution. Therefore, the original techniques (8) and (9) are at least disputable (for small portfolios of reinsurers).

Accepting the previous arguments, we prefer the expected shortfall methods given by (13) amongst all models introduced in Section 3 mainly due to their stability over various N (compare with the results for (10) and (12) in Figure 1). Alternatively, the value-at-risk based calculation (10) and (12) might be preferred if one implements some mechanism to determine a minimal value of required capital (see e.g. (8)) and is aware of possible higher volatility of required capital.

Furthermore, Tables 2 – 4 contain $RC_{0.995}$ for the portfolio of N similar risk expositions with equal credit rates and $LGD_i = 1/N$ for $i = 1,\ldots,N$ ($N = 1,\ 5,\ 10$) for all methods introduced in Section 3. The RC variation within a particular cell amongst all tables describes the diversification benefit to an insurer if one increases the number of counterparties (reinsurers). Note that the variation in a particular column is typically much larger than within a particular cell amongst all tables. Apparently, there is not much diversification benefit if the insurer goes from one to, e.g., five reinsurers, especially for very highly rated reinsurance companies. Obviously, it is much more important to have a high-quality reinsurer rather than ceding business to a number of low-quality reinsurers. Diversification benefit is material for the speculative class and low important for lower credit risk classes, which is the same conclusion as in [6].

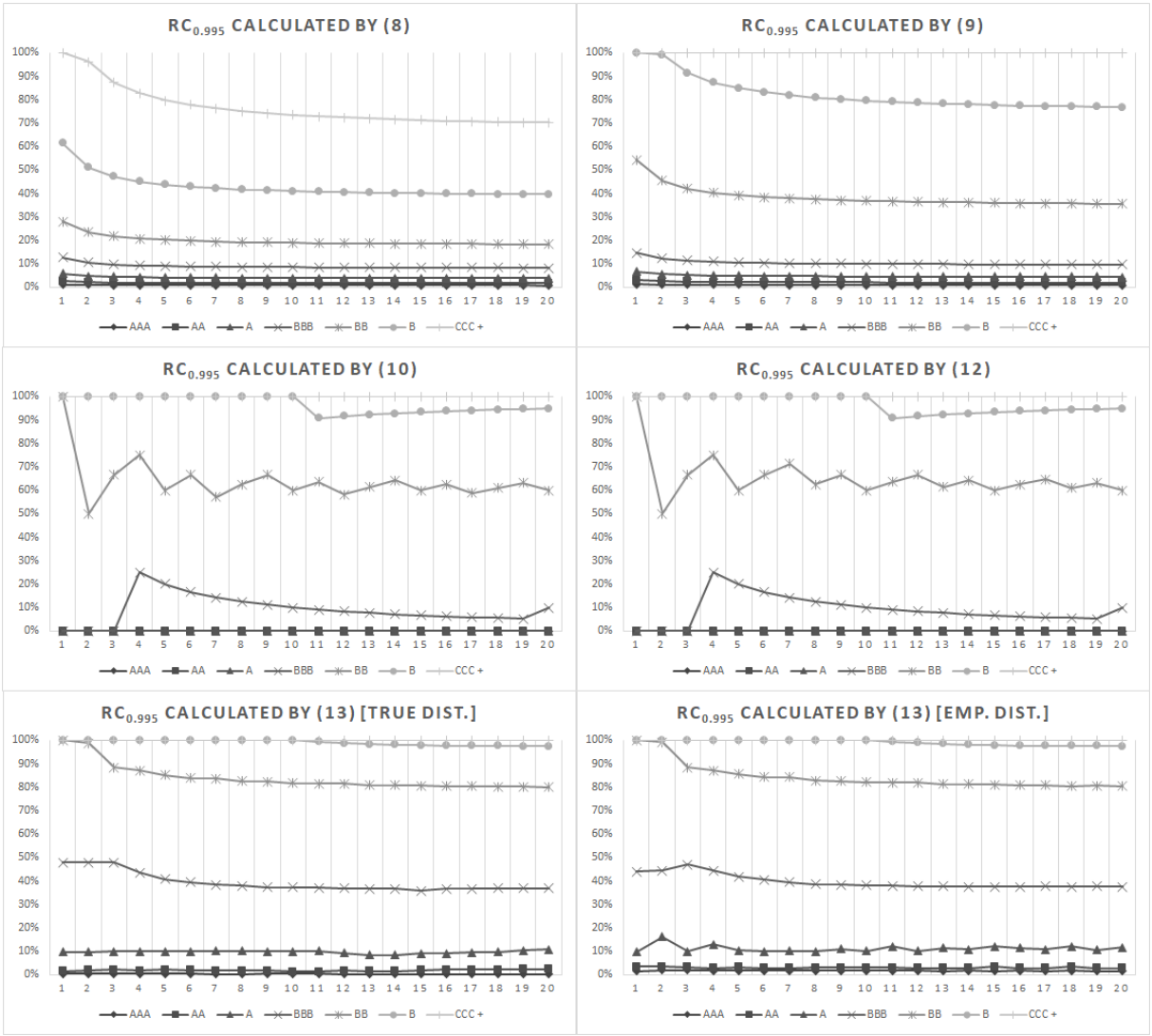


Figure 1 Required capital $RC_{0.995}$ calculated by methods in Section 3

5 Conclusion

The paper dealt with the construction of required capital to cover the default risk in portfolios with a smaller number of heterogeneous counterparties. We surveyed various methods for determining counterparty default risk of reinsurance based on the common shock principle, proposed modifications of these methods and compared them by means of a numerical study. In particular, we investigated the portfolio of N similar risk expositions (reinsurers)

with equal credit rates and equal loss contributions. The numerical results confirm that the suggested modifications of the widely accepted approach introduced in [9], which was also implemented within the Solvency II framework (see [4] and [8]), might be preferred. In any case, the recommended methods confirm the fact that a single high-quality counterparty is more effective in this context than a number of low-quality counterparties.

Method	(8)	(9)	(10)	(12)	(13): True dist.	(13): Emp. dist.
AAA	1.2%	1.3%	0.0%	0.0%	0.5%	1.5%
AA	2.6%	3.0%	0.0%	0.0%	1.5%	3.5%
A	5.8%	6.7%	0.0%	0.0%	9.6%	9.6%
BBB	12.6%	14.7%	0.0%	0.0%	48.0%	43.9%
BB	28.0%	54.4%	100.0%	100.0%	100.0%	100.0%
B	61.4%	100.0%	100.0%	100.0%	100.0%	100.0%
CCC +	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 2 $RC_{0.995}$ for the portfolio of one risk exposition with $LGD_1 = 1$

Method	(8)	(9)	(10)	(12)	(13): True dist.	(13): Emp. dist.
AAA	0.8%	1.0%	0.0%	0.0%	0.4%	1.7%
AA	1.9%	2.2%	0.0%	0.0%	2.1%	3.1%
A	4.2%	4.8%	0.0%	0.0%	10.0%	10.4%
BBB	9.1%	10.6%	20.0%	20.0%	40.8%	41.8%
BB	20.2%	39.2%	60.0%	60.0%	85.3%	85.7%
B	43.7%	84.9%	100.0%	100.0%	100.0%	100.0%
CCC +	79.8%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 3 $RC_{0.995}$ for the portfolio of 5 similar risk expositions with equal credit rates and $LGD_i = 0.2$ for all i

Method	(8)	(9)	(10)	(12)	(13): True dist.	(13): Emp. dist.
AAA	0.8%	0.9%	0.0%	0.0%	0.4%	1.6%
AA	1.7%	2.0%	0.0%	0.0%	1.3%	2.9%
A	3.9%	4.5%	0.0%	0.0%	9.9%	10.1%
BBB	8.5%	10.0%	10.0%	10.0%	37.2%	38.1%
BB	19.0%	36.8%	60.0%	60.0%	81.8%	82.1%
B	41.0%	79.6%	100.0%	100.0%	100.0%	100.0%
CCC +	73.5%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 4 $RC_{0.995}$ for the portfolio of 10 similar risk expositions with equal credit rates and $LGD_i = 0.1$ for all i

Acknowledgements

This research was supported by the grant 17–00676S provided by the Grant Agency of the Czech Republic.

References

[1] Cipra, T. (2010). *Financial and Insurance Formulas*. New York: Springer.
[2] Cipra, T. & Hendrych, R (2017). Systemic risk in financial risk regulation. *Czech Journal of Economics and Finance*, *67*, 15–38.
[3] Hürlimann, W (2008). Solvency II reinsurance counterparty default risk. *Life & Pension*, *December 2008*, 39–44.
[4] QIS5 (2011). *EIOPA Quantitative Impact Study 5* (available at: https://eiopa.europa.eu/publications/reports/ qis5_report_final.pdf).
[5] Quick, R. & Weglarz, D (2012). Reinsurance default risk. *Gen Re*, *October 2012* (available at: http://media.genre.com/documents/InsuranceIssues201210-en.pdf).
[6] Sachs, R. (2009). Reinsurance credit risk. *Munich Re Group Technical Report*, *2009*, 267–285.
[7] Sandström, A. (2011). *Handbook of Solvency for Actuaries and Risk Managers. Theory and Practice*. New York: Chapman and Hall / CRC Press.
[8] Solvency II (2009). *Directive on the taking-up and pursuit of the business of insurance and reinsurance. Directive 2009/138/EC of the European Parliament and of the Council 2009*.
[9] Ter Berg, P. (2008). Portfolio modelling of counterparty reinsurance default risk. *Life & Pension*, *April 2008*, 29–33.
[10] Vasicek, O. A. (2002). The distribution of loan portfolio value. *Risk*, *15*, 160–162.

Matrix games with uncertain entries: A robust approach

Robert Hlavatý[1], Helena Brožová[2]

Abstract. In classical game theory, a conflict of two opponents can be modelled as an equilibrium-based matrix game. We assume a conflict of two non-cooperative antagonistic opponents with a finite number of strategies with zero-sum or constant sum pay-offs. In the same time, we suppose that the elements of the payoff matrix describing the game are not fixed and are allowed to change within a specified interval.
Supposing that some of the elements of the payoff matrix are uncertain, it is evident that this would influence the utilities of both players at the same time and moreover, such entropy of the model would eventually influence the position of equilibria or its very existence.
We propose a modelling approach that allows one to find a solution of the game with either pure or mixed strategies of opponents with the guaranteed payoffs under the assumption that a specified number of unspecified entries would attain different values than expected. The chosen robust approach is presented briefly as well as the necessary circumstances of matrix game solutions. Our novelty approach follows and is accompanied by an explanatory example in the end of the paper.

Keywords: Game theory, matrix games, robust approach, equilibria, uncertainty

JEL Classification: C70
AMS Classification: 91A05, 90C08, 90C31

1 Introduction

Decision-making process is heavily influenced by the uncertainty in the decision-making situation and the inaccuracy of the data. From this perspective, it is possible to analyze the results of models of game theory, if we assume possible changes in payoffs. Payoff changes affect the choice of the players' best strategies (pure or mixed) and moreover, a change that is advantageous to one of the players may be disadvantageous to the other. There are different approaches for treating uncertainty in operations research. A stochastic approach is probably the one that comes to one's mind when uncertainty is mentioned. This approach is probabilistic-based. A fuzzy approach is also available when it comes to calculations with inexact parameters. We want to take advantage of robust optimization tools that are strictly deterministic and rather set-based than probabilistic [4] to deal with the goals of this paper. Since the matrix game can be reformulated as a linear optimization problem, we are going to treat the uncertainty here in a sense of robust linear optimization.

The parameter uncertainty in the linear optimization has been dealt with in many ways since the pioneering work of Dantzig [8] that discusses the very matter of linear programming under uncertainty. The first extensive work on the matters of robust optimization is considered by Soyster [15]. In his paper, he proposes the reformulation of a linear program into the new program that contains the aspects of uncertainty and is able to provide robust solutions that are "protected" against the adverse changes in parameters. The Soyster's approach preserves linearity, it tends to be, however, very over-conservative meaning that it seeks to protect the model results at all costs from all possible uncertainty influences which also brings a significant losses in objective function value.

Much later, Ben-Tal and Nemirovski [3] proposed an approach where the conservativeness of the protection can be controlled by additional parameters, while this formulation leads to conic quadratic programming that is non-linear itself. Later on, a novelty approach was introduced by Bertsimas and Sim [5] that allows controlling the level of conservatism in the solution in a way that leads once again to a linear optimization model that can even be applied to discrete optimization problems. This approach was further extended by Büsing and D'Andreagiovanni [6] and subsequently elaborated in detail in 2014 by the same authors (Büsing and D'Andreagiovanni [7]). Finally, it is worth to mention an extensive work of Ben-Tal, El Ghaoui and Nemirovski [2] describing the issues of Robust optimization as a whole under the different perspectives.

The following paragraph discusses the results achieved on uncertainty in game theory by other authors. The probabilistic approach is presented by Pun and Wong [13] wherein it is shown how one can possibly apply their approach specifically in the insurance, which is modelled as a game here. Fuzzy approach is also very popular in

[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

game theory. Namely in matrix-type of games, the pareto-optimal solutions with fuzzy payoffs are discussed by Aggarwal, Chandra and Mehra [1] or in a bit more specific fashion by Dutta and Gupta [9] who discuss the Nash equilibrium in the games with trapezoidal fuzzy payoffs. Similar matters are presented for bi-matrix games under the fuzzy entries by Roy and Mula [14] while Li [12] is concerned with the triangular fuzzy entries and also an effective approach to computation of the problem. It is discussed by the same author [11], how one is able to treat the matrix games using deterministic approach, specifically the interval arithmetic while he also demonstrates the means of reaching the results of the game by linear programming. While in Li's [11] approach the outcome of the game is once again interval data, we propose an approach seeking robust solutions such that input parameters are considered to deviate within an assumed interval.

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 matrix game and its reformulation as a linear program.

2.1 Robust programming introduction

In the following 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 [5] also described in our previous paper [10]. 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_i, c_j (or at least some of them) are not precisely defined. For purposes of this article, let us consider the uncertainty to be present within the set of coefficients a_{ij} from now on. This leads to a new problem:

$$\begin{aligned} & max \sum_{j=1}^{n} c_j x_j \\ & \text{s.t.} \\ & \sum_{j=1}^{n} \left(a_{ij} + \delta_{ij}^{a}\right) x_j \leq b_i \,, i = 1, \dots, m \\ & x_j \geq 0 \,, j = 1, \dots, n \end{aligned} \tag{2}$$

The (2) represents the reformulation of (1) with uncertain coefficients a_{ij}. The uncertainty is expressed using deviations for any coefficients if need be. We assume any deviation δ_{ij}^{a} to be any real nonzero number. It was illustrated by Ben-Tal, El Ghaoui and Nemirovski [2] 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. Before a construction of the robust counterpart of (1), Bertsimas and Sim [5] 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 one 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):

$$\begin{gathered} max \sum_{j=1}^{n} c_j x_j \\ \text{s.t.} \\ \sum_{j=1}^{n} a_{ij} x_j + \Gamma_i z_i + \sum_{j \in Q_i} q_{ij} \leq b_i \, , i = 1, \dots, m \\ z_i + q_{ij} \geq \delta_{ij}^{a} x_j \, , i = 1, \dots, m \, , \forall j \in Q_i \\ z_i \geq 0 \, , i = 1, \dots, m \\ q_{ij} \geq 0 \, , i = 1, \dots, m \, , \forall j \in Q_i \\ x_j \geq 0 \, , j = 1, \dots, n \end{gathered} \tag{3}$$

where the parameter $\Gamma_i, 0 \leq \Gamma_i \leq |Q_i|, i = 1, \dots, m$, controls the protection against uncertainty in the constraint i, q_{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 Q_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 Matrix game

Let us consider the basic type of a game called *matrix game* or *normal-form game*. This game describes a zero-sum conflict of two players H_1 and H_2 each one with the finite number of strategies $R_i, i \in \{1, \dots, m\}$ and $S_j, j \in \{1, \dots, n\}$ respectively. A matrix $\boldsymbol{A} \in \mathbb{R}^{mxn}$ with entries a_{ij} is called payoff matrix. All entries represent payments from H_2 to H_1 and since we deal with the zero-sum game, it is possible to construct an equivalent payoff matrix for opposite direction payments $\bar{\boldsymbol{A}} \in \mathbb{R}^{mxn}$ with entries $\bar{a}_{ij} = -a_{ij}, \forall i \in \{1, \dots, m\}, j \in \{1, \dots, n\}$.

The game and its solution is based on von-Neumann's minimax theorem meaning that both players seek to minimize the maximum payoff of the other player. Thus, it is necessary to find vectors $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ of optimal strategies portfolio for both players respectively. Componentwise $\boldsymbol{r}^* = (r_1, r_2, \dots, r_m)$ and $\boldsymbol{s}^* = (s_1, s_2, \dots, s_n)$ show the optimal usage of any strategy R_i (S_j) for both players. Necessarily $\|r^*\|_1 = 1$ and $\|s^*\|_1 = 1$ for any case. We seek such pair of vectors $\boldsymbol{r}^*, \boldsymbol{s}^*$ that $\max_{\boldsymbol{r}} \min_{\boldsymbol{s}} \boldsymbol{r^T A s} = \min_{\boldsymbol{s}} \max_{\boldsymbol{r}} \boldsymbol{r^T A s}$ in the game. One is able to determine the solution easily in a sense of pure strategies. This can be achieved by applying the minmax (or maxmin) approach directly in the matrix to determine which particular strategy (or strategies in case there is more equilibria in the game) optimize the payoffs. In general, there does not have to be any equilibrium such that usage of a unique strategy optimizes the payoffs of both players. In that case, it is necessary to determine all components of $\boldsymbol{r}^*, \boldsymbol{s}^*$ to obtain frequencies with which the strategies should be used to maximize outcomes while minimizing losses.

Any *matrix game* can be reformulated as a linear program based on the minimax assumption while normalizing the optimal strategy vectors to 1. The minimax objective formulation is, however, not a linear expression and it is generally NP-hard in terms of solution complexity. By introducing a new variable w as the first player's outcome, it is possible to bound the expected payoff of the player as $\boldsymbol{r^T a_j} \geq \boldsymbol{w}, \boldsymbol{w} = (w, .., w), \forall j \in \{1, \dots, n\}$ while we maximize w. Similarly, this can be done for the second player by bounding the objective from above by a variable w in the following way: $\boldsymbol{a_i s} \leq \boldsymbol{w}, \forall i \in \{1, \dots, m\}$ while minimizing w. Considering the both substitutions, it is possible to formulate a pair of linear programs

$$\begin{gathered} \max w \\ \text{s.t.} \\ \boldsymbol{r^T a_j} \geq \boldsymbol{w}, j = 1, \dots, n \\ \sum_{i=1}^{m} r_i = 1 \\ \boldsymbol{x} \geq 0 \end{gathered} \qquad \begin{gathered} \min w \\ \text{s.t.} \\ \boldsymbol{a_i s} \leq \boldsymbol{w}, i = 1, .., m \\ \sum_{j=1}^{n} s_j = 1 \\ \boldsymbol{y} \geq 0 \end{gathered} \tag{4}$$

These programs can be transformed into pair of dual linear programs.

$$\begin{gathered} w^* = min \sum_{i=1}^{m} x_i \\ \text{s.t.} \end{gathered} \qquad \begin{gathered} w^* = max \sum_{i=1}^{m} y_i \\ \text{s.t.} \end{gathered} \tag{5}$$

$$\boldsymbol{x^T\hat{a}_j \geq 1}, j = 1, \dots, n \qquad \boldsymbol{\hat{a}_i y \leq 1}, i = 1, .., m$$
$$\boldsymbol{x \geq 0} \qquad \boldsymbol{y \geq 0}$$

From the strong duality theorem, for these particular programs it holds that the optimal solutions are equal. This value lies between upper and lower value of the matrix game $\underline{w} \leq w^* \leq \overline{w}$.

Finally, before the linear programs can be practically solved, one must deal with the fact that the both models will only yield desired results when all entries a_{ij} are positive. The entries in the real life situations can also be non-positive. This can be easily dealt with by transforming all entries additively by some α that will shift all values positive in the following fashion: $\alpha = \left|\min_{i,j} a_{ij}\right| + c$, with a positive constant c and new entries $\hat{a}_{ij}$ are $\hat{a}_{ij} = a_{ij} + \alpha, \forall i, j$.

For a matrix game, the following theorem is true, which shows that if players play optimal strategies, both will achieve the best possible result.

The Minimax Theorem.

For every finite two-person zero-sum game,

(1) there is a number w^*, $\underline{w} \leq w^* \leq \overline{w}$ called the value of the game,

(2) there is a mixed strategy for Player H_1 such that his average gain is at least w^* no matter what H_2 does, and

(3) there is a mixed strategy for Player H_2 such that his average loss is at most w^* no matter what H_1 does.

3 Results

The assumption that some of the input data of the matrix game are not precisely known, it leads a decision maker to finding robust solution. The concept of building a robust counterpart was briefly described in the chapter 2.1 and it allows a player to embed uncertainties into the left-hand side matrix of the system $\hat{A}\boldsymbol{x} \geq 1$ or $\boldsymbol{y^T}\hat{A} \leq 1$. Considering the dual linear programs defining the game from both views, the only changeable parameters lie within the left-hand side coefficients. This also make sense in terms of the original game representation by the payoff table where the only input data were actually the payoffs a_{ij}.

3.1 Robust formulation of the matrix game

We will take advantage of the formulation of robust counterpart in (3) and specify it for the particular problems of the matrix game. Note that the indices i, j represent rows and columns of A_{ij} whereas in the following model the representation is transposed:

$$\underline{w} = min \sum_{i=1}^{m} x_i \qquad \overline{w} = min \sum_{i=1}^{m} y_j$$
s.t. s.t.
$$\sum_{i=1}^{m} x_i \hat{a}_{ij} - \Gamma_j z_j - \sum_{i \in Q_j} q_{ji} \geq 1, j = 1, \dots, n \qquad \sum_{i=1}^{m} y_j \hat{a}_{ij} + \Gamma_i z_i + \sum_{i \in Q_j} q_{ji} \leq 1, i = 1, \dots, m \quad (6)$$
$$z_j + q_{ji} \geq \delta_{ji}^a x_i, \forall i \in Q_j, j = 1, \dots, n \qquad z_i + q_{ji} \geq \delta_{ji}^a y_j, \forall i \in Q_j, i = 1, \dots, m$$
$$\boldsymbol{z \geq 0} \qquad \boldsymbol{z \geq 0}$$
$$q_{ji} \geq 0, \forall i \in Q, j = 1, \dots, n \qquad q_{ji} \geq 0, \forall i \in Q, i = 1, \dots, m$$
$$x_i \geq 0, i = 1, \dots, m \qquad y_j \geq 0, j = 1, \dots, n$$

where the parameter $\Gamma_j, 0 \leq \Gamma_j \leq \left|Q_j\right|$ controls the protection against uncertainty in a chosen constraint, i.e. how many coefficients a_{ji} are expected to deviate from its deterministic value in the constraint j, i resp.

Because these models do not create the pair of dual linear programs, it is possible to suppose, that $\underline{w} \leq \overline{w}$ which means that the worst outcome of the first player according to the first player selection of strategy is lower than the first player's outcome according to the second player's selection of strategy.

3.2 Practical example

To illustrate the usage of the proposed robust approach, we will demonstrate it on the following small-scale matrix game with an arbitrarily chosen input. Two players H_1, H_2 have available their strategy sets $R = \{R_1, R_2, R_3\}$ and

$S = \{S_1, S_2, S_3, S_4\}$ with the payoff matrix $\boldsymbol{A}$ and the maximum possible deviations of some of payoffs in matrix $\boldsymbol{\Delta}$ chosen arbitrarily as follows:

$$A = \begin{pmatrix} 1 & 2 & 5 & 7 \\ 4 & 2 & 1 & 8 \\ 2 & 5 & 3 & 6 \end{pmatrix} \qquad \Delta = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 4 \\ 1 & 1 & 2 & 5 \end{pmatrix} \tag{7}$$

Looking for the saddle point, the initial estimation of payoffs gives the boundary on the feasible value of the game, the lower value is $\underline{w} = 2$ and the upper value is $\overline{w} = 4$. The optimal mixed strategy of the players are $\boldsymbol{r}^{*T} = (0.304;\ 0.478;\ 0.218)$ and $\boldsymbol{s}^{*T} = (0.522;\ 0.087;\ 0.391;\ 0)$ and the value of the game is $w = 2.652$. When the real payoffs decrease and reaches the lowest values, the lower and upper value of the game is $\underline{w} = \overline{w} = 1$ and the game has a saddle point, and in opposite case, when the payoffs increase to their upper limits, the lower and upper value of the game are $\underline{w} = 3$ and $\overline{w} = 5$ (*Table 1*).

	S1	S2	S3	S4	Min
R1	1	2	5	7	1
R2	1	2	1	4	1
R3	1	4	1	1	1
Max	1	4	5	7	

The lowest payoffs

	S1	S2	S3	S4	Min
R1	1	2	5	7	1
R2	4	2	1	8	1
R3	2	5	3	6	2
Max	4	5	5	8	

Initial estimation of payoffs

	S1	S2	S3	S4	Min
R1	1	2	5	7	1
R2	7	2	1	12	1
R3	3	6	5	11	3
Max	7	6	5	12	

The highest payoffs

Table 1 Payoff matrices

Suppose now that at most four payoffs can change ($\sum \Gamma_j \leq 4$ and $\sum \Gamma_i \leq 4$). The robust model of the first player is

$$\underline{w} = min \sum_{i=1}^{m} x_i \text{ subject to:} \quad \begin{gathered} \begin{pmatrix} 1 & 4 & 2 \\ 2 & 2 & 5 \\ 5 & 1 & 3 \\ 7 & 8 & 6 \end{pmatrix} \boldsymbol{x} - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \boldsymbol{z} - \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \\ q_{41} & q_{42} & q_{43} \end{pmatrix} \geq \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \\ - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{x} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} z_1 + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_{11} \\ q_{12} \\ q_{13} \end{pmatrix} \geq 0 \\ - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{x} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} z_2 + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_{21} \\ q_{22} \\ q_{23} \end{pmatrix} \geq 0 \\ - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \boldsymbol{x} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} z_3 + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_{31} \\ q_{32} \\ q_{33} \end{pmatrix} \geq 0 \\ - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{pmatrix} \boldsymbol{x} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} z_4 + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_{41} \\ q_{42} \\ q_{43} \end{pmatrix} \geq 0 \\ \boldsymbol{z} \geq 0, \boldsymbol{y} \geq 0, \boldsymbol{Q} \geq 0 \end{gathered} \tag{8}$$

This model shows the worst-case strategy for assumed deviations with the minimal guaranteed objective function. The optimal strategy for the first player is $\boldsymbol{r}^{*T} = (0.158;\ 0.210;\ 0.632)$ and the value of the game is $w = 1.632$. This outcome is better than the outcome that can be obtained in case of decreasing of all payoffs (respective lower value of game is $\underline{w} = 1$). The robust form of model for the second player is following

$$\overline{w} = max \sum_{j=1}^{n} z_j \text{ subject to:} \quad \begin{gathered} \begin{pmatrix} 1 & 2 & 5 & 7 \\ 4 & 2 & 1 & 8 \\ 2 & 5 & 3 & 6 \end{pmatrix} \boldsymbol{y} + \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} \boldsymbol{z} + \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{pmatrix} \leq \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ - \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \boldsymbol{y} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} z_1 + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_{11} \\ q_{12} \\ q_{13} \\ q_{14} \end{pmatrix} \geq 0 \\ - \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \boldsymbol{y} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} z_2 + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_{21} \\ q_{22} \\ q_{23} \\ q_{24} \end{pmatrix} \geq 0 \\ - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix} \boldsymbol{y} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} z_3 + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_{31} \\ q_{32} \\ q_{33} \\ q_{34} \end{pmatrix} \geq 0 \\ \boldsymbol{z} \geq 0, \boldsymbol{y} \geq 0, \boldsymbol{Q} \geq 0 \end{gathered} \tag{9}$$

Similarly, the optimal strategy for the second player is $\boldsymbol{s}^{*\boldsymbol{T}} = (0.5;\ 0;\ 0.5;\ 0)$ and the value of the game is $w = 4$. And again, this outcome is better than the outcome that can be obtained in case of increasing of all payoffs (respective upper value of the game is $\overline{w} = 5$).

Analysis of a matrix game with uncertain evaluation of payoffs can be done using the robust approach. The value received by solving the robust model for the first player can be seen as the lower bound of outcomes of the players, which is higher than lower value of the game with lowest payoffs. In contrary, the value gained by solving the robust model for the second player is the upper bound of the players' outcomes, which is lower than the upper value of the game with the highest payoffs. Of course, these bounds depend on the assumptions of the supposed changes of payoffs. Advantage of this way of game analysis is that we receive the tighter interval of possible outcomes.

4 Conclusion

We have explored the possibilities of applying Γ-robustness approach in the matrix games of two players. It was discussed in the introduction part that the different techniques could be used for evaluation when uncertain entries of the payoff matrix are present. Each technique brings the decision maker slightly different point of view on the results achieved by computations, i.e. probabilistic approach leads to probabilistic outcomes, fuzzy entries will lead to fuzzy character of the result etc. We applied an approach that is solely deterministic and generates deterministic results, too. Using this approach not only allows the decision maker to control a level of protection against uncertainty by the Γ parameter but it also occurred, as shown in our example, that the evaluation by robust approach leads to tighter bounds on the value of the game. It remains to our future work to explore how does this particular robust approach generally influences an existence of equilibria of the game.

References

[1] Aggarwal, A., Chandra, S. and Mehra, A.: Solving Matrix Games with I-fuzzy Payoffs: Pareto-optimal Security Strategies Approach. *Fuzzy Information and Engineering* **6** (2014), 167-192.

[2] Ben-Tal, A., El Ghaoui A.L. and Nemirovski A.: *Robust optimization*. Princeton University Press, 2009.

[3] Ben-Tal, A. and Nemirovski, A.: Robust solutions of uncertain linear programs. *Operations Research Letters* **25** (1999), 1-13.

[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.: New 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.: A New Theoretical Framework for Robust Optimization under Multi-band Uncertainty. In: *Operations Research Proceedings*, 2014, 115-121.

[8] Dantzig, G.B.: Linear programming under uncertainty. *Management Science* **1** (1955), 197-206.

[9] Dutta, B. and Gupta, S.K.: On Nash Equilibrium Strategy of Two-person Zero-sum Games with Trapezoidal Fuzzy Payoffs. *Fuzzy Information and Engineering* **6** (2014), 299-314.

[10] Hlavatý, R. and Brožová, H.: Robust optimization approach in transportation problem. In *35th Conference Mathematical Methods in Economics 2017* (pp. 225-230).

[11] Li, D.: Linear Programming Approach to Solve Interval-valued Matrix Games. *Omega* **39** (2011), 655-666.

[12] Li, D.: A Fast Approach to Compute Fuzzy Values of Matrix Games with Payoffs of Triangular Fuzzy Numbers. *European Journal of Operational Research* **223** (2012), 421-429.

[13] Pun, C.S. and Wong, H.Y.: Robust non-zero sum stochastic differential reinsurance game. *Insurance: Mathematics and Economics* **68** (2016), 169-177.

[14] Roy, S.K. and Mula, P.: Bi-matrix game in bifuzzy environment. *Journal of Uncertainty Analysis and Applications* **1**(11) (2013).

[15] Soyster, A.L.: Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research,* **21**, (1973), 1154-1157.

Interval Estimation of Quadratic Variation

Vladimír Holý [1], Ondřej Sokol [2]

Abstract. Financial processes such as stock prices, exchange rates and commodity prices are observed as intervals because of the bid-ask spread and discreteness of prices. We investigate the effects of this interval uncertainty on quadratic variation estimation. We assume that the price process can be decomposed into a continuous component and a discrete component with a finite number of jumps. The jump component can further be decomposed into a small-jump component containing jumps smaller than a given treshold and a large-jump component containing jumps larger than this treshold. We prove that quadratic variation is unbounded from above under interval uncertainty. The quadratic variation for the continuous as well as small-jump component is also not identifiable as the upper bound is infinite. The quadratic variation for the large-jump component is, however, bounded from above and therefore partially identifiable. We prove the related theorems and illustrate the estimation of quadratic variation for the large-jump component in a simulation study.

Keywords: High-Frequency Data, Quadratic Variation, Interval Analysis.

JEL classification: C58
AMS classification: 62M09, 91G70

1 Introduction

An important aspect of quantitative finace is a volatility analysis of stock prices, exchange rates and commodity prices [1, 8]. Intraday volatility is typically measured by quadratic variation of the price process with continuous time. For this purpose, high-frequency data are utilized. An important characteristic of high-frequency data is that the price process is not observed directly and precisely. The indirectness is because instead of actual price process, only bid and ask prices are observed. The imprecision is because the prices are quoted and recorded only with a given precision.

Traditionally, the bid-ask spread and discreteness of prices are modeled by the market microstructure noise. In this setting, the price process (called the efficient price) is contaminated by an additive noise with zero expected value. The presence of this noise in prices causes significant bias of realized variance – an estimator of quadratic variation [7]. In the literature, there are several methods robust to this noise. Namely, the two-scale estimator [17], the realized kernels [2], the pre-averaging estimator [9] and the least square estimator [14].

We take a different approach as we do not capture the bid-ask spread and discreteness of prices by an additive noise but rather by interval uncertainty. This means that we assume that the exact values are not available, only the bounds are observable. This problem is known as partial identification [12]. As only lower and upper bounds of the price process are available, the goal is to compute lower and upper bounds of certain statistics. However, even some of the basic statistics are not easy to estimate in this setting. The statistics studied in the literature include sample variance [4, 5, 15], t-ratio [3], entropy [10, 16] and higher moments [11]. The summary of interval uncertainty methods can be found in [6] and [13].

The rest of the paper is structured as follows. In Section 2, we define quadratic variation and prove some basic theorems related to its decomposition. In Section 3, we present interval uncertainty and prove theorems related to the bounds of quadratic variation when the process is observed only as intervals. In Section 4, we illustrate finite-sample estimation of quadratic variation for the large jumps under interval uncertainty in a simulation study. We conclude the paper in Section 5.

2 Theoretical Framework

Let us consider the proccess P_t with continuous time $t \geq 0$. We assume this process can be decomposed into the *continuous component* P_t^C with continuous values and the *jump component* P_t^J with a finite number of discrete

[1]University of Economics, Prague, Winston Churchill Square 1938/4, 130 67 Prague 3, Czech Republic, vladimir.holy@vse.cz
[2]University of Economics, Prague, Winston Churchill Square 1938/4, 130 67 Prague 3, Czech Republic, ondrej.sokol@vse.cz

jumps, i.e. $P_t = P_t^C + P_t^J$. Formally, the jump component is defined as

$$P_t^J = \sum_{i:0\le S_i\le t} J_i, \qquad t \ge 0,$$

where J_i are non-zero random variables with non-equal random times S_i. We further decompose the jump component into the *small-jump component* P_t^{SJ} containing only jumps smaller or equal to a given threshold κ in absolute value and the *large-jump component* P_t^{LJ} containing only jumps larger than $\kappa > 0$ in absolute value, i.e. $P_t^J = P_t^{SJ} + P_t^{LJ}$. Formally, the small-jump component and the large-jump component are defined as

$$P_t^{SJ} = \sum_{i:0\le S_i\le t, |J_i|\le\kappa} J_i, \qquad P_t^{LJ} = \sum_{i:0\le S_i\le t, |J_i|>\kappa} J_i, \qquad t \ge 0,$$

where J_i are non-zero random variables with non-equal random times S_i. For a given $\kappa > 0$, the price component can then be decomposed into $P_t = P_t^C + P_t^{SJ} + P_t^{LJ}$.

Our main focus is the quadratic variation. Without a loss of generality, we limit ourselves to the time interval $[0, 1]$. Let us consider a sampling of the process P_t at discrete times $0 = T_0 < T_1 < \ldots < T_n = 1$. *Quadratic variation* is then given by

$$QV_{[0,1]} = \operatorname*{plim}_{\Delta^n\to 0} \sum_{i=1}^{n} \left(P_{T_i} - P_{T_{i-1}}\right)^2,$$

where plim denotes the limit in probability and $\Delta^n = \max\{T_1 - T_0, T_2 - T_1, \ldots, T_n - T_{n-1}\}$ is the maximal lag between the observations. We denote $QV_{[0,1]}^C$, $QV_{[0,1]}^{SJ}$ and $QV_{[0,1]}^{LJ}$ quadratic variation for the continuous component P_t^C, small-jump component P_t^{SJ} and large-jump component P_t^{LJ} respectively.

Theorem 1. *Quadratic variation can be decomposed into* $QV_{[0,1]} = QV_{[0,1]}^C + QV_{[0,1]}^J$.

Proof. We decompose quadratic variation as

$$\begin{aligned}
QV_{[0,1]} &= \operatorname*{plim}_{\Delta^n\to 0} \sum_{i=1}^{n} \left(P_{T_i}^C - P_{T_{i-1}}^C + P_{T_i}^J - P_{T_{i-1}}^J\right)^2 \\
&= \operatorname*{plim}_{\Delta^n\to 0} \left(\sum_{i=1}^{n} \left(P_{T_i}^C - P_{T_{i-1}}^C\right)^2 + \sum_{i=1}^{n} \left(P_{T_i}^J + P_{T_{i-1}}^J\right)^2 + 2\sum_{i=1}^{n} \left(P_{T_i}^C - P_{T_{i-1}}^C\right)\left(P_{T_i}^J - P_{T_{i-1}}^J\right) \right) \\
&= QV_{[0,1]}^C + QV_{[0,1]}^J + 2 \operatorname*{plim}_{\Delta^n\to 0} \sum_{i=1}^{n} \left(P_{T_i}^C - P_{T_{i-1}}^C\right)\left(P_{T_i}^J - P_{T_{i-1}}^J\right).
\end{aligned}$$

As $P_{T_i}^J - P_{T_{i-1}}^J$ is nonzero only in a finite number of cases, we have

$$\operatorname*{plim}_{\Delta^n\to 0} \sum_{i=1}^{n} \left(P_{T_i}^C - P_{T_{i-1}}^C\right)\left(P_{T_i}^J - P_{T_{i-1}}^J\right) = \sum_{i:0\le S_i\le 1} \operatorname*{plim}_{\Delta^n\to 0} \left(P_{S_i}^C - P_{S_i-\Delta^n}^C\right)\left(P_{S_i}^J - P_{S_i-\Delta^n}^J\right).$$

Because

$$\operatorname*{plim}_{\Delta^n\to 0} \left(P_{S_i}^C - P_{S_i-\Delta^n}^C\right) = 0 \quad \text{and} \quad \operatorname*{plim}_{\Delta^n\to 0} \left(P_{S_i}^J - P_{S_i-\Delta^n}^J\right) = J_i,$$

we have

$$\operatorname*{plim}_{\Delta^n\to 0} \left(P_{S_i}^C - P_{S_i-\Delta^n}^C\right)\left(P_{S_i}^J - P_{S_i-\Delta^n}^J\right) = 0,$$

and therefore,

$$\sum_{i:0\le S_i\le 1} \operatorname*{plim}_{\Delta^n\to 0} \left(P_{S_i}^C - P_{S_i-\Delta^n}^C\right)\left(P_{S_i}^J - P_{S_i-\Delta^n}^J\right) = 0.$$

□

Theorem 2. *Quadratic variation for the jump component is* $QV_{[0,1]}^J = \sum_{i:0\le S_i\le 1} J_i^2$. *Quadratic variation* $QV_{[0,1]}^J$ *can be decomposed into* $QV_{[0,1]}^J = QV_{[0,1]}^{SJ} + QV_{[0,1]}^{LJ}$.

Proof. As $P^J_{T_i} - P^J_{T_{i-1}}$ is nonzero only in a finite number of cases, we have

$$QV^J_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \sum_{i=1}^{n} \left(P^J_{T_i} - P^J_{T_{i-1}}\right)^2 = \sum_{i:0\le S_i\le 1} \operatorname*{plim}_{\Delta^n \to 0} \left(P^J_{S_i} - P^J_{S_i-\Delta^n}\right)^2.$$

Because

$$\operatorname*{plim}_{\Delta^n \to 0} \left(P^J_{S_i} - P^J_{S_i-\Delta^n}\right) = J_i,$$

we have

$$\operatorname*{plim}_{\Delta^n \to 0} \left(P^J_{S_i} - P^J_{S_i-\Delta^n}\right)^2 = J_i^2,$$

and therefore,

$$\sum_{i:0\le S_i\le 1} \operatorname*{plim}_{\Delta^n \to 0} \left(P^J_{S_i} - P^J_{S_i-\Delta^n}\right)^2 = \sum_{i:0\le S_i\le 1} J_i^2.$$

The second proposition is a direct implication of this. □

3 Interval Approach

Let us consider we do not observe the process P_t, $t \ge 0$ but rather a collection of intervals $[\underline{P}_t, \overline{P}_t]$ such that it is guaranteed that $P_t \in [\underline{P}_t, \overline{P}_t]$, $t \ge 0$. This setup, in which the true values are not observable but the bounds are available, is known as partial identification. Due to our weak assumptions, the only information we can infer about any statistic from the observable intervals $[\underline{P}_t, \overline{P}_t]$, $t \ge 0$ is its lower and upper bound. In the case of quadratic variation, $QV_{[0,1]} \in [\underline{QV}_{[0,1]}, \overline{QV}_{[0,1]}]$, where $\underline{QV}_{[0,1]}$ and $\overline{QV}_{[0,1]}$ is the lower and upper bound of the form

$$\underline{QV}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \min\Big\{ \sum_{i=1}^{n} \left(\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}\right)^2 : \underline{P}_{T_i} \le \tilde{P}_{T_i} \le \overline{P}_{T_i},\ i = 0, \dots, n \Big\},$$

$$\overline{QV}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \max\Big\{ \sum_{i=1}^{n} \left(\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}\right)^2 : \underline{P}_{T_i} \le \tilde{P}_{T_i} \le \overline{P}_{T_i},\ i = 0, \dots, n \Big\}.$$

We assume a process $\tilde{P}_t$ can be decomposed into the continuous component $\tilde{P}^C_t$, the small-jump component $\tilde{P}^{SJ}_t$ and the large-jump component $\tilde{P}^{LJ}_t$ just like the original process P_t. Similarly, we define $\underline{QV}^C_{[0,1]}$, $\overline{QV}^C_{[0,1]}$, $\underline{QV}^{SJ}_{[0,1]}$, $\overline{QV}^{SJ}_{[0,1]}$, $\underline{QV}^{LJ}_{[0,1]}$, $\overline{QV}^{LJ}_{[0,1]}$ as the minimal/maximal quadratic variation for processes $\tilde{P}^C_t$/$\tilde{P}^{SJ}_t$/$\tilde{P}^{LJ}_t$ over all possible processes $\tilde{P}_t \in [\underline{P}_t, \overline{P}_t]$. For simplification, we assume all intervals $[\underline{P}_t, \overline{P}_t]$ have constant width ω, i.e. $\omega = \overline{P}_t - \underline{P}_t$ for all $t \ge 0$. The threshold separating small and large jumps is then set to $\kappa = 2\omega$.

In the following theorems, we investigate properties of quadratic variation under interval uncertainty. We find that quadratic variation is unbounded from above. Quadratic variation for both the continuous component and the small-jump component is also unbounded from above. Quadratic variation for the large-jump component is, however, bounded from above and therefore partially identifiable.

Theorem 3. *Quadratic variation is unbounded from above, i.e.* $\overline{QV}_{[0,1]} = \infty$.

Proof. The upper bound of quadratic variation is given by

$$\overline{QV}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \max\Big\{ \sum_{i=1}^{n} \left(\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}\right)^2 : \underline{P}_{T_i} \le \tilde{P}_{T_i} \le \overline{P}_{T_i},\ i = 0, \dots, n \Big\}.$$

Let us consider process $\hat{P}_{T_i}$ defined for $i = 0$ as $\hat{P}_{T_0} = \underline{P}_{T_0}$ and for $i = 1, \dots, n$ as $\hat{P}_{T_i} = \underline{P}_{T_i}$ if $|\hat{P}_{T_{i-1}} - \underline{P}_{T_i}| > |\hat{P}_{T_{i-1}} - \overline{P}_{T_i}|$ or $\hat{P}_{T_i} = \overline{P}_{T_i}$ else. For any two consecutive values, we have $|\hat{P}_{T_i} - \hat{P}_{T_{i-1}}| \ge \frac{\omega}{2}$. Therefore, we have

$$\frac{\omega^2}{4} n \le \sum_{i=1}^{n} \left(\hat{P}_{T_i} - \hat{P}_{T_{i-1}}\right)^2 \le \max\Big\{ \sum_{i=1}^{n} \left(\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}\right)^2 : \underline{P}_{T_i} \le \tilde{P}_{T_i} \le \overline{P}_{T_i},\ i = 0, \dots, n \Big\}.$$

For $\Delta^n \to 0$, we have $n \to \infty$ and quadratic variation diverges to infinity. □

Theorem 4. *Quadratic variation for the continuous component as well as the small-jump component is unbounded from above, i.e.* $\overline{QV}^{C}_{[0,1]} = \infty$ *and* $\overline{QV}^{SJ}_{[0,1]} = \infty$ *respectively.*

Proof. The upper bound of quadratic variation for the continuous component is given by

$$\overline{QV}^{C}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \max\Big\{ \sum_{i=1}^{n} \Big(\tilde{P}^{C}_{T_i} - \tilde{P}^{C}_{T_{i-1}} \Big)^2 : \ \underline{P}_{T_i} \leq \tilde{P}_{T_i} \leq \overline{P}_{T_i}, \ i = 0, \ldots, n \Big\}.$$

while the upper bound of quadratic variation for the small-jump component is given by

$$\overline{QV}^{SJ}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \max\Big\{ \sum_{i=1}^{n} \Big(\tilde{P}^{SJ}_{T_i} - \tilde{P}^{SJ}_{T_{i-1}} \Big)^2 : \ \underline{P}_{T_i} \leq \tilde{P}_{T_i} \leq \overline{P}_{T_i}, \ i = 0, \ldots, n \Big\},$$

where $\tilde{P}^{SJ}_t = \sum_{j:0 \leq S^S_j \leq t, |J^S_j| > 2\omega} J^S_j$, $t \geq 0$ is the small-jump component with jumps $\tilde{J}^S_j$ and jump times $\tilde{S}^S_j$ corresponding to a process $\tilde{P}_t$. As there are only finite number of values that can be attributed to jumps, a nontrivial time interval $[a, b] \subset [0, 1]$ with $\min\{ \operatorname*{plim}_{\Delta^n \to 0} |\overline{P}_t - \underline{P}_{t-\Delta^n}|, \operatorname*{plim}_{\Delta^n \to 0} |\underline{P}_t - \overline{P}_{t-\Delta^n}| \} \leq 2\omega$ for each $t \in (a, b]$ exists. This interval cannot contain any large jumps, i.e. $\overline{QV}^{LJ}_{[a,b]} = 0$, and the upper bound of quadratic variation for continuous component is indistinguishable from the small-jump component, i.e. $\overline{QV}_{[a,b]} = \overline{QV}^{C}_{[a,b]} = \overline{QV}^{SJ}_{[a,b]}$. From Theorem 3, we have $\overline{QV}_{[a,b]} = \infty$. Finally, we have $\overline{QV}^{C}_{[0,1]} \geq \overline{QV}^{C}_{[a,b]} = \infty$ and $\overline{QV}^{SJ}_{[0,1]} \geq \overline{QV}^{SJ}_{[a,b]} = \infty$. □

Theorem 5. *The lower and upper bounds of quadratic variation for the large-jump component are finite and respectively given by*

$$\underline{QV}^{LJ}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \sum_{i=1}^{n} \underline{\chi}^2_{T_i, T_{i-1}} \mathbb{I}_{\{\underline{\chi}_{T_i, T_{i-1}} > 2\omega\}},$$

$$\overline{QV}^{LJ}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \sum_{i=1}^{n} \overline{\chi}^2_{T_i, T_{i-1}} \mathbb{I}_{\{\overline{\chi}_{T_i, T_{i-1}} > 2\omega\}},$$

where $\mathbb{I}$ *denotes the indicator function and*

$$\underline{\chi}_{T_i, T_{i-1}} = \min\{ |\overline{P}_{T_i} - \underline{P}_{T_{i-1}}|, |\underline{P}_{T_i} - \overline{P}_{T_{i-1}}| \},$$

$$\overline{\chi}_{T_i, T_{i-1}} = \max\{ |\overline{P}_{T_i} - \underline{P}_{T_{i-1}}|, |\underline{P}_{T_i} - \overline{P}_{T_{i-1}}| \}.$$

Proof. The lower and upper bounds of quadratic variation for the large-jump component are defined as

$$\underline{QV}^{LJ}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \min\Big\{ \sum_{i=1}^{n} \Big(\tilde{P}^{LJ}_{T_i} - \tilde{P}^{LJ}_{T_{i-1}} \Big)^2 : \ \underline{P}_{T_i} \leq \tilde{P}_{T_i} \leq \overline{P}_{T_i}, \ i = 0, \ldots, n \Big\},$$

$$\overline{QV}^{LJ}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \max\Big\{ \sum_{i=1}^{n} \Big(\tilde{P}^{LJ}_{T_i} - \tilde{P}^{LJ}_{T_{i-1}} \Big)^2 : \ \underline{P}_{T_i} \leq \tilde{P}_{T_i} \leq \overline{P}_{T_i}, \ i = 0, \ldots, n \Big\},$$

where $\tilde{P}^{LJ}_t = \sum_{j:0 \leq S^L_j \leq t, |J^L_j| > 2\omega} J^L_j$, $t \geq 0$ is the large-jump component with jumps $\tilde{J}^L_j$ and jump times $\tilde{S}^L_j$ corresponding to a process $\tilde{P}_t$. An absolute difference between two consecutive values $\operatorname*{plim}_{\Delta^n \to 0} |\tilde{P}^{LJ}_{T_i} - \tilde{P}^{LJ}_{T_i - \Delta^n}|$ can either be $|\tilde{J}^L_j| > 2\omega$ or zero from the definition of the large-jump component. If a large jump $\tilde{J}^L_j$ occurs at time T_i, we have $2\omega < \operatorname*{plim}_{\Delta^n \to 0} |\tilde{P}_{T_i} - \tilde{P}_{T_i - \Delta^n}| = \operatorname*{plim}_{\Delta^n \to 0} |\tilde{P}^{LJ}_{T_i} - \tilde{P}^{LJ}_{T_i - \Delta^n}| = |\tilde{J}^L_j|$ as the absolute difference between two consecutive values in limit is zero for both the continuous and small-jump component. Therefore, we can ignore $\operatorname*{plim}_{\Delta^n \to 0} |\tilde{P}_{T_i} - \tilde{P}_{T_i - \Delta^n}| \leq 2\omega$ as they do not correspond to a large jump. On the other hand, $\operatorname*{plim}_{\Delta^n \to 0} |\tilde{P}_{T_i} - \tilde{P}_{T_i - \Delta^n}| > 2\omega$ may correspond to a large jump and must be included. We can replace $\operatorname*{plim}_{\Delta^n \to 0} |\tilde{P}^{LJ}_{T_i} - \tilde{P}^{LJ}_{T_i - \Delta^n}|$ with $\operatorname*{plim}_{\Delta^n \to 0} |\tilde{P}_{T_i} - \tilde{P}_{T_i - \Delta^n}|$ as we have a finite number of $\operatorname*{plim}_{\Delta^n \to 0} |\tilde{P}_{T_i} - \tilde{P}_{T_i - \Delta^n}| > 2\omega$. The lower and upper bounds are then given by

$$\underline{QV}^{LJ}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \min\Big\{ \sum_{i=1}^{n} \Big(\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}} \Big)^2 \mathbb{I}_{\{|\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}| > 2\omega\}} : \ \underline{P}_{T_i} \leq \tilde{P}_{T_i} \leq \overline{P}_{T_i}, \ i = 0, \ldots, n \Big\},$$

$$\overline{QV}^{LJ}_{[0,1]} = \operatorname*{plim}_{\Delta^n \to 0} \max\Big\{ \sum_{i=1}^{n} \Big(\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}} \Big)^2 \mathbb{I}_{\{|\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}| > 2\omega\}} : \ \underline{P}_{T_i} \leq \tilde{P}_{T_i} \leq \overline{P}_{T_i}, \ i = 0, \ldots, n \Big\}.$$

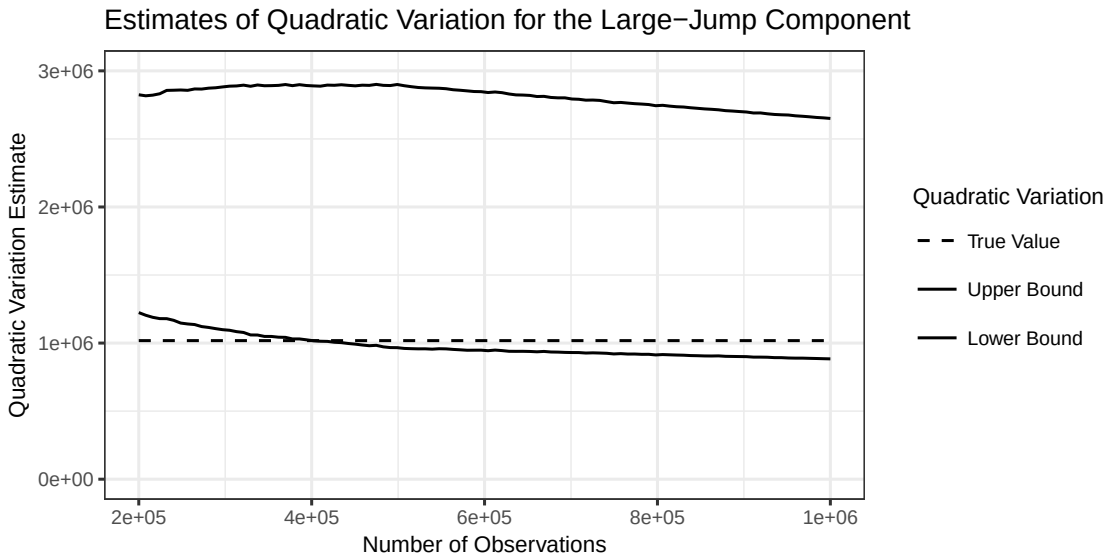


Figure 1 A simulation of lower and upper bounds of quadratic variation for the large-jump component.

As two jumps cannot occur at the same time, we have

$$\underline{QV}_{[0,1]}^{LJ} = \operatorname*{plim}_{\Delta^n \to 0} \sum_{i=1}^{n} \min\left\{\left(\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}\right)^2 \mathbb{I}_{\{|\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}| > 2\omega\}} : \underline{P}_{T_i} \leq \tilde{P}_{T_i} \leq \overline{P}_{T_i}\right\},$$

$$\overline{QV}_{[0,1]}^{LJ} = \operatorname*{plim}_{\Delta^n \to 0} \sum_{i=1}^{n} \max\left\{\left(\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}\right)^2 \mathbb{I}_{\{|\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}| > 2\omega\}} : \underline{P}_{T_i} \leq \tilde{P}_{T_i} \leq \overline{P}_{T_i}\right\}.$$

Finally, we can restrict the minimization and maximization to extreme points and get the desired expressions with $\underline{\chi}_{T_i,T_{i-1}}$ and $\overline{\chi}_{T_i,T_{i-1}}$ respectively. Both bounds are finite because $\operatorname*{plim}_{\Delta^n \to 0} |\tilde{P}_{T_i} - \tilde{P}_{T_{i-1}}| > 2\omega$ can occur only at finite number of times. □

4 Simulation Study

In a simulation study, we illustrate the finite-sample properties of estimation of quadratic variation for the large-jump component. We simulate 1 000 000 observations as the sum of the continuous and jump component. The continuous component is simulated as the Wiener process with zero mean and unit standard deviation. The jump component contains 10 000 jumps with values generated from the normal distribution with zero mean and standard deviation equal to 10 and with times generated according to the exponential distribution.

The finite-sample counterpart of quadratic variation is called the realized variance. Simulated interval estimates of quadratic variation (i.e. interval realized variances) for various numbers of observations are presented in Figure 1. We can see that for a smaller number of observations, the interval estimate is not precise as it omits some jumps. For a larger number of observations, the interval estimate converges and the bounds contain the true value of quadratic variation for the large-jump component.

5 Conclusion

We deal with the estimation of quadratic variation under interval uncertainty. First, we prove that quadratic variation can be decomposed into a continuous and jump component. Second, we show that quadratic variation for the jump component is the sum of squared jump values. Third, we prove that quadratic variation is unbounded from above when the process is not directly observed and we have only intervals available. Fourth, we also find that quadratic variation for the continuous and jump component containing small jumps is unbounded from above under interval uncertainty. Fifth, we prove that quadratic variation for the jump component containing large jumps is bounded from above and present lower and upper tight bounds. Finally, we illustrate finite-sample estimation of quadratic variation for the large-jump component in a simulation study. Our results can be utilized in a volatility analysis of financial processes containing large jumps.

Acknowledgements

This paper was supported by the grant No. F4/58/2017 of the Internal Grant Agency of University of Economics, Prague.

References

[1] Aït-Sahalia, Y., and Jacod, J.: *High-Frequency Financial Econometrics*. Princeton University Press, Princeton, 2014.

[2] 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.

[3] Černý, M., and Hladík, M.: The Complexity of Computation and Approximation of the t-Ratio Over One-Dimensional Interval Data. *Computational Statistics & Data Analysis& Data Analysis* **80** (2014), 26–43.

[4] Černý, M., and Sokol, O.: Interval Data and Sample Variance: A Study of an Efficiently Computable Case. In: *Proceedings of the 33th International Conference Mathematical Methods in Economics*. Plzeň, 99–105.

[5] Ferson, S., Ginzburg, L., Kreinovich, V., Longpré, L., and Aviles, M.: Exact Bounds on Finite Populations of Interval Data. *Reliable Computing* **11** (2005), 207–233.

[6] Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W. L., and Ginzburg, L.: *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*. Technical report, 2007.

[7] Hansen, P. R., and Lunde, A.: Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics* **24** (2006), 127–161.

[8] Hautsch, N.: *Econometrics of Financial High-Frequency Data*. Springer, 2011.

[9] 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.

[10] Kreinovich, V.: Maximum Entropy and Interval Computations. *Reliable Computing* **2** (1996), 63–79.

[11] Kreinovich, V., and Longpre, L.: *Computing Higher Central Moments for Interval Data*. Technical report, 2004.

[12] Manski, C. F.: *Partial Identification of Probability Distributions*. Springer, New York, 2003.

[13] Nguyen, H. T., Kreinovich, V., Wu, B., and Xiang, G.: *Computing Statistics under Interval and Fuzzy Uncertainty*. Springer, Berlin, Heidelberg, 2012.

[14] Nolte, I., and Voev, V.: Least Squares Inference on Integrated Volatility and the Relationship Between Efficient Prices and Noise. *Journal of Business & Economic Statistics* **30** (2012), 94–108.

[15] Sokol, O.: How to Prove Polynomiality of Computation of Upper Bound Considering Random Intervals? In: *Proceedings of the 34th International Conference Mathematical Methods in Economics*. Liberec, 779–785.

[16] Xiang, G., Ceberio, M., and Kreinovich, V.: Computing Population Variance and Entropy under Interval Uncertainty: Linear-Time Algorithms. *Reliable Computing* **13** (2007), 467–488.

[17] Zhang, L., Mykland, P. A., and Aït-Sahalia, Y.: A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. *Journal of the American Statistical Association* **100** (2005), 1394–1411.

How to tell if a hockey player performs well (enough)

Jakub Houdek [1], Ondřej Sokol [2]

Abstract. In this paper, we deal with the analysis of the hockey player's performance. An algorithm for data preprocessing is introduced as well as a linear regression model which allows us to estimate hockey players' skills.
Data is obtained using `R` and `nhlscrapr` package which extracts data from NHL's Real Time Scoring System. The result is a special +/- statistic which is adjusted to a single player's performance, thus we call it adjusted plus-minus (APM).
The goal of this papers is to decide whether a selected NHL player performs well (or at least well enough) and then use the information to estimate team's performance. Therefore, a logistic regression model which aggregates player performances is created. This model allows us to estimate which team is supposed to be well performing. We compare the estimates with the real results. The results are promising.

Keywords: Ice hockey, linear regression, logistic regression, r.

JEL classification: C44
AMS classification: 90C15

1 Motivation for analysis

The hockey world is simply put goal-oriented. This actually has two meanings. First, every hockey team in NHL has an ultimate goal of winning the Stanley Cup, which is the trophy that the league awards to the best NHL team. Second, if you want to win this trophy, you have to win a lot of matches throughout the season that consists of 82 regular season games and of a various number of playoff games (the number of playoff matches depends on the series' lengths). In hockey world there is one simple saying that leads to a win – *just score one more goal than the opposing team*.

Every team tries to do anything to win. This includes trading for highly effective players, selecting best available players at the entry draft before every season, offering team's players yoga lessons (inspiration from NFL, football league) or creating an analytics department to provide a precise evaluation of player skill and abilities.

There have been several attempts at creating models which could give the teams' managements better information about players. Precise evaluation of players is crucial in assembling of the team especially nowadays in a salary cap era in which each team has limited space for players' payrolls. A successful team cannot afford to spend money on ineffective players.

In this particular paper, a model, which was described by Brian Macdonald [2], is followed and then back-tested using the recent data. We also analyze problems of the proposed model. The model developed for back-testing is centered around the assumption that Macdonald's model provides us with values that describe players' skills that are isolated from other players' successes as we cannot accurately measure individual players' performances as ice hockey is a team-based sport.

2 Acquiring data from NHL

Datasets were acquired using `nhlscrapr` package [4] via `R` environment [3].

The initial dataset consists of a large number of data matrices in which one matrix stands for one match. These matrices contain data about every single event that has happened during the match on the ice. Every event also gives us information about the time of occurrence, who was on the ice during the event and the type of the event. Original data matrices also contain other information such as the date of the match but these information are not needed for the purpose of this analysis.

This analysis was done for seasons beginning in years 2011, 2012, 2013 and 2014 separately. What Macdonald did in [2] is that he joined several previous seasons in one and then performed his analysis.

3 Creating the model

Complete model can be divided into two parts. First part is ranking players based on their skills and then aggregating their skills and performing a logistic regression to back-test the overall skill of every team. Using this model

[1]Department of Econometrics, University of Economics, Prague, houdek@forsaga.net
[2]Department of Econometrics, University of Economics, Prague, ondrej.sokol@vse.cz

we can model the chance of the team of making playoffs.

3.1 Linear regression model for acquiring players' skills

Macdonald [2] describes the model for evaluating players' skills as following:

$$y_i^{ES} = \beta_0 + \sum_{j=1}^{J} h_{ij}\beta_j + \sum_{j=1}^{J} a_{ij}\delta_j + e_i, \tag{1}$$

where y_i^{ES} is the number of goals that would be scored in favor of the opposing team, if players present on ice during the event i were playing for full 60 minutes of the game, $i = 1 \dots N$ is indexing the events on the ice for all events that have happened on ice, β_0 defines the home-ice advantage, h_{ij} denotes j-th home player's presence on the ice during the event i, J is the total number of players that have occurred on ice during the selected match, β_j denotes the offensive skill of the player j, a_{ij} denotes j-th away player's presence on the ice during the event i, δ_j denotes the defensive skill of the player j and e_i is the error term.

This equation describes the estimates of the number of goals scored if selected players are present on ice. What we do, is that we apply this equation to every line change that has occurred. During the analysis we also eliminate events which are not important for us – we only want to know how long a certain line combination was on the ice and whether a goal was scored.

When applied to a whole match, we can determine what players were more effective than the others based on players' β and δ coefficients. What is troublesome is the fact that players in lines are not usually switched during the match. This leads to estimates that have very high standard deviation.

In order to minimize this problem with higher standard deviations in players' skills coefficients β and δ, we need to apply this model to all matches (or as many as possible).

It is important to note that in this model we actually need to 'turn the sides'. Players who are currently in the possession of the puck and are attacking are also defending in a sense. Let's imagine this situation – one team has a line so dominant that it scores a lot of goals, plays really well and does not get any goals scored upon. Therefore we need to look at it from the point of view of the opposing team.

Therefore, we need to adjust the model. This final model can be described as

$$\underset{2N\times 1}{\begin{bmatrix} y^{ES} \\ 0 \end{bmatrix}} = \underset{2N\times 1}{\beta_0} + \underset{2N\times 2J}{\begin{bmatrix} HA \\ AH \end{bmatrix}} \underset{2J\times 1}{\begin{bmatrix} \beta \\ \delta \end{bmatrix}} + \underset{2N\times 1}{\epsilon}, \tag{2}$$

where the left hand side represents a vector where first half of elements are the goals scored by attacking teams and the second half of elements are the goals scored by the defending team. Right hand side represents a big matrix which submatrices represent players present on the ice during the shifts.

Remark. For every shift that occurred on the ice, we find players that have appeared on the ice. We determine who was attacking and who was defending. Then we are ought to determine the length of the shift. Afterward, we determine whether a goal was scored. If a goal was scored, we evaluate the number of goals that would be scored if the line would be playing the whole match – full 60 minutes, e.g. $3600 y_i^{ES}/t_i \in \mathbb{R}_{>0}$. If a goal wasn't scored, then $3600 y_i^{ES}/t_i = 0$ as the number of goals scored is 0. Afterward, we need to create a second equation where the sides are switched and represent the game from the point of view of the other team. We apply this to all matches we desire to be analyzed.

3.2 Conclusion of Macdonalnd's model

Using this model we obtain estimates of β and δ. However, these estimates need to be transformed in a following way:

$$OPM = \hat{\beta}, \tag{3}$$

$$DPM = -\hat{\delta}, \tag{4}$$

$$APM = OPM + DPM, \tag{5}$$

where OPM is an estimated offensive skill of a selected player, DPM is an estimated defensive skill of the same player and APM is the total skill of this player.

This way we get APM, which the estimate that is essential for evaluating player's performance on ice and thus essential for the following model, where we back-test Macdonald's model and at the same time try to make a prediction whether a team can advance to playoffs.

3.3 Back-testing of Macdonalnd's model

Premise of Macdonalnd's model is that it should isolate players' skills, therefore aggregating their skills should give us the overall strength all teams.

Creating a ranking model for NHL teams

Aggregation was simply done using a weighted sum as following:

$$APMW_{team} = \sum_{i \in team} APM_i \frac{\bar{t}_i}{3600}, \quad team \in NHL \tag{6}$$

where NHL is a set of all NHL teams.

This way was created the complete ranking of teams participating in NHL.

Usage of the ranking model

NHL is one of many major hockey leagues where the winner of the season is not determined by points but using playoff bracket. Every team's goal is to enter the playoffs. So instead of a creation of points predicting model, playoff model was created.

This model is described as

$$PLAYOFF_{team} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 APMW_{team} + u_{team})}}, \quad team \in NHL \tag{7}$$

Intuitively, this equation described odds of entering playoffs based on teams' $APMW$ value.

4 Evaluating results

4.1 Evaluation of player rankings

Although the total count of players appearing during the selected season is not unbearable for researching whether the player's APM is adequate, it is rather impossible for one human to be able to actually know whether the player is performing well. This problem occurs mostly when trying to evaluate less famous players. These players can be performing well but there are not many sources documenting it aside from regular or advanced statistics provided by NHL itself or any statistics oriented fan websites.

Sufficient way to inspect whether the ranking 'makes sense' is the usage of histograms of statistics defined in previous paragraphs.

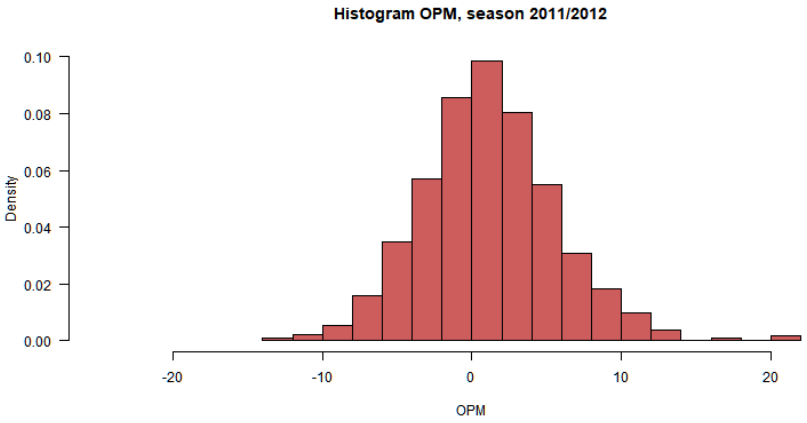


Figure 1 OPM histogram

In the case of OPM histogram (see Figure 4.1) we can see that the histogram resembles normal distribution.

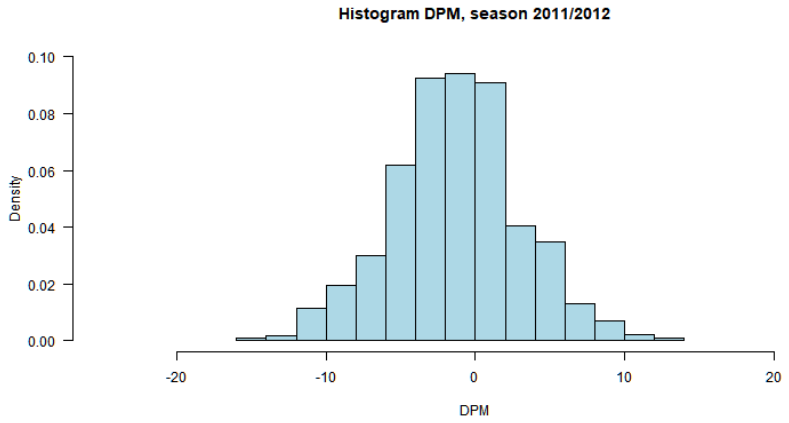


Figure 2 DPM histogram

On the other hand, DPM histogram does not show that obvious similarity to normal distribution but can still be considered adequate.

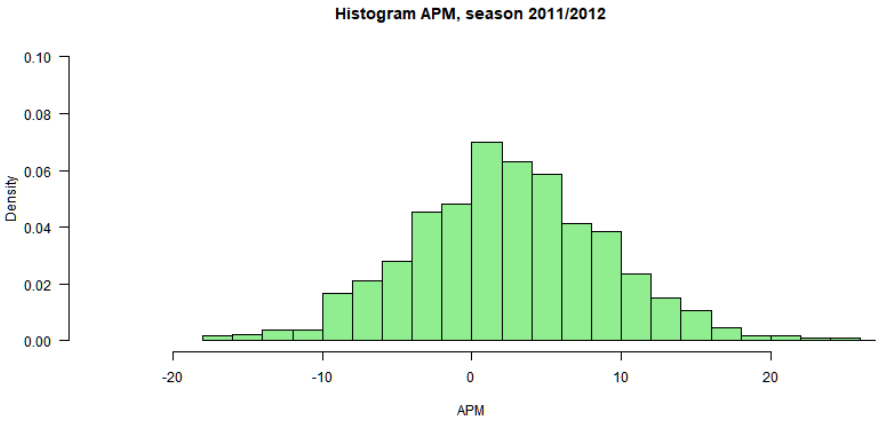


Figure 3 APM histogram

Deciding on distribution of the histogram for APM (see Figure 4.1) values is more complicated as it is derived from equation 5, which is a sum of two random variables that are usually negatively correlated as seen in Figure 4. The size displays the total time spent on ice. The negative correlation can be explained through the inability of an average player to perform well in both defending and attacking. It is quite common to describe players as defensive or offensive players, therefore the regression line confirms this common description. However, there are only several players that could be considered purely offensive or purely defensive.

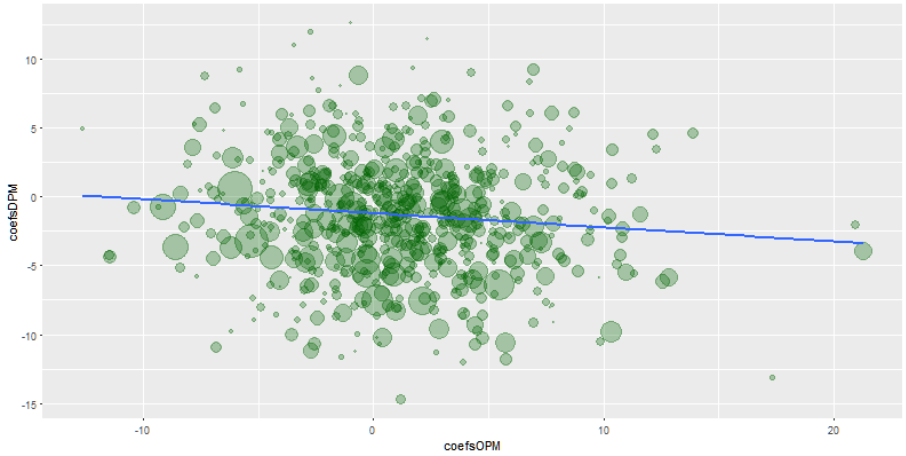


Figure 4 Negative correlation between OPM and DPM

4.2 Back-testing evaluation

In order to determine the successfulness of back-testing, it is needed to examine, whether the logistic model is able to determine the probability of a team entering the playoffs. This can be seen in Figure 5.

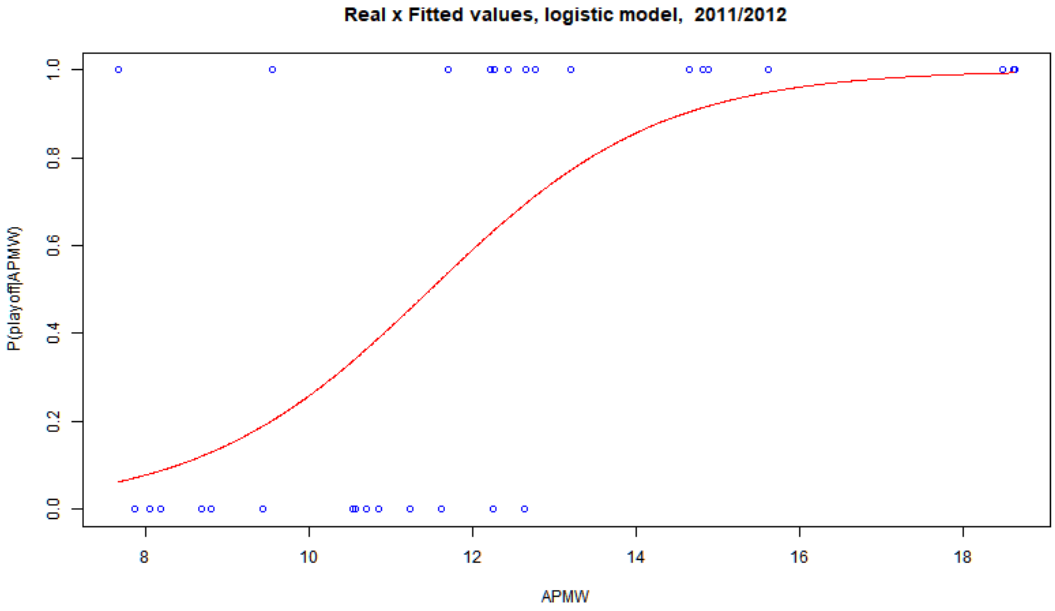


Figure 5 Plotting of a logistic curve and comparing it with real results

We can see three distinguishable groups of teams. First group consists of teams that have not made playoffs and their $APMW$ is rather low. On the other side, we can see teams that have made it to the playoffs and have rather high $APMW$ value. The third group represents the teams that have rather average $APMW$ and may or may not make the playoff during this season. Based on examining the seasons' results this is not surprising as some of the teams tend to be average and their playoff appearance might be more or less based on luck and other teams' results.

5 Conclusions and considerations

While Macdonald's model provides us with a manual how to rank players, it is still not that efficient. Macdonald in his later work[1] adds goaltenders and the effect of special teams. This provides us an additional information about the performance of all teams. While the results might seem accurate, they are still incomplete.

One of the flaws is that some players tend to play with the same players for the entire season. And in some cases they play with the same player or players for their entire career. Sedin twins can be a good example, which was addressed by Macdonald himself. This leads to an inability to identify the skill of one of the twins, or in less extreme cases, inability to distinguish skills of players that have a line 'chemistry'.

The so called chemistry is one of the flaws that is not covered in any of Macdonald's models. Chemistry can cause players to suddenly overperform, which temporarily inflates players' skills and therefore it makes the estimation of players' skills more difficult.

While the common sense tells us to join seasons and create a larger dataset with more events, that should lead to minimizing variance in player skills, this might not be the best way to eliminate the high variance problem. Player skills tend to be inconsistent between seasons. In some cases we can observe players having overall under average career, but they have one season that is much better that others.

Originally designed method for back-testing provided us with results that seem to be sufficient and tells us, that Macdonald's model provides us with informations that are valuable but still are not as accurate as they can be due to reasons described in this paper.

Future possible back-testing method for players' performances can include game theory models, where we take in consideration an assumption, that coaches of NHL teams are selecting lines to ice based on their best possible decision - *select the line that is the most efficient against the opponent's line*, and thus converge into a goal differential which in this case would be represented by the *price of the game*.

References

[1] Brian Macdonald. An improved adjusted plus-minus statistic for nhl players. In *Proceedings of the MIT Sloan Sports Analytics Conference*, 03 2011.

[2] Brian Macdonald. A regression-based adjusted plus-minus statistic for nhl players. *Journal of Quantitative Analysis in Sports*, 7, 07 2011.

[3] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[4] A.C. Thomas and Samuel L. Ventura. *nhlscrapr: Compiling the NHL Real Time Scoring System Database for easy use in R*, 2017. R package version 1.8.1.

SOM with Diffusion Modelling in Stock Market Analysis

Radek Hrebik[1], Jaromir Kukal[2], Josef Jablonsky[3]

Abstract. Traditional self organised map (SOM) is learned by Kohonen paradigm. Novel model of self organisation is based on diffusion modelling in continuous space which is a good approximation of endorphins propagation in real brain. Therefore the structure of system is described by neuron coordinates instead of neighborhood relationship in traditional SOM. Neuron activation using diffusion process is discussed. Novel diffusive learning algorithm is based on this activation mentioned above. Paper is focused on market analysis via Kohonen and diffusion learning. Kohonen SOM is used as referential method. Using logarithmic differences of stock market data is preferred as data preprocessing.

Keywords: self organization, Kohonen map, diffusion learning, SOM, stock market.

JEL classification: C63
AMS classification: 68T05

1 Introduction

There are many approaches how to perform modelling of self organisation. They can be directly inspired by anatomy and physiology of neuronal system or rather by other ideas which are easy to realize. Our research is inspired by pudding model of atom in physics [1, 2], where the nucleus of atoms are supposed as points (raisins) in the electron continuum (pudding). In the case of self organisation we will place individual neurons instead of atom nucleus into the continuum which would transfer the information in the system. The second inspiration is strongly connected with brain physiology study about slow signal propagation in central nervous system.

The SOM can be directly applied to the analysis of time series of stock prices. The logarithmic differences i.e. daily yields are evaluated for given stock first. Then we can define the state of the stock as its history in previous days. These states are real vectors of x-dimensions which are in mutual relationships and the relationship structural is useful for the study of stock market evaluation. This approach will be applied both to individual stocks and to their system.

2 Pudding Model of SOM

Our model of self-organised map is based on specific assumptions:

- Finite number of neurons is placed in fix positions like raisins in a pudding.
- The neurons are surrounded by unconstrained continuum as analogy of the pudding base.
- Neuron interconnections are omitted.
- The pattern set stays outside the pudding and only sequentially activates individual neurons.
- Neuron activities generate concentration profile of substrate in the pudding.
- Substrate concentration influences the learning rates of individual neurons.
- The learning process changes weights as neuron properties.

The *pudding model* description begins with remembering of basic facts. Let $m, n, H \in \mathbb{N}$ be number of patterns, pattern dimensionality and number of SOM neurons [3]. 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 fixed positions of individual neurons in continuum are $\mathbf{p}_i \in \mathbb{R}^N$ for $i = 1, ..., H$ and reflects the topology of SOM [4] which is subject of network design. The diffusion process in continuum can be easily expressed using matrix $\mathbf{D} \in (\mathbb{R}_0^+)^{H \times H}$ of mutual distances $d_{i,j} = \|\mathbf{p}_i - \mathbf{p}_j\|_2$. Therefore the resulting SOM is invariant to translation and rotation of its structure. Let $\Delta t > 0$ be learning period and the diffusion in continuum will be studied only in discrete time $t_k = k \cdot \Delta t$, where $k \in \mathbb{N}_0$. The result of SOM learning is the system of weights [5] $\mathbf{w}_i \in \mathbb{R}^n$, where $i = 1, ..., H$ of course. We begin with random weights setting $\mathbf{w}_i(0)$. The weights evolve during learning process and their values in time t_q are denoted as $\mathbf{w}_i(q)$, where $q \in \mathbb{N}_0$. The *pudding model* is based on substrate concentrations in neurons and given time. Being prepared to SOM learning we have to study the concentration profile first.

[1]CTU in Prague, FNSPE, Brehova 7, Prague 1, Czech Republic, Radek.Hrebik@seznam.cz
[2]CTU in Prague, FNSPE, Brehova 7, Prague 1, Czech Republic, Jaromir.Kukal@fjfi.cvut.cz
[3]University of Economics, FIS, W. Churchill Square 4, Prague 3, Czech Republic, Josef.Jablonsky@vse.cz

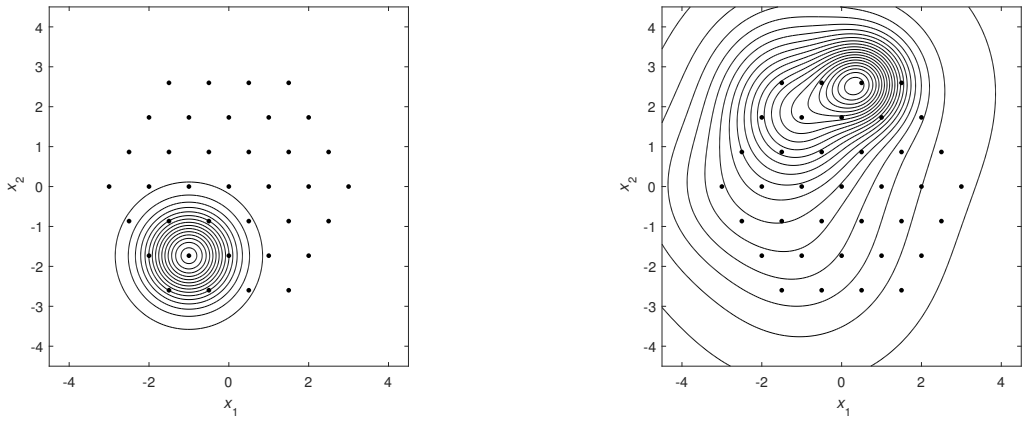


Figure 1: Concentration profile after single (left) and complete (right) activation ($N = 2, a = 1, b = 1/10, H = 37, q = 100$)

2.1 Single Activation

The pudding SOM learning is based on the activation of single neuron. We will study j-th neuron which is supposed to be active in time t_k. Therefore, formally $j = \varphi_k$. The concentration profile in $\mathbb{R}^N$ is depicted on the left part of figure 1 for $N = 2$. But it is not necessary to study the substrate concentration in any point. The learning is based only on the concentration in neuron points. The concentration in time t_q is

$$\mathrm{c}(\mathbf{y}, \mathbf{p}_j, t_q) = \frac{1}{(4\pi D(t_q - t_k))^{N/2}} \cdot \exp\left(-\frac{\|\mathbf{y} - \mathbf{p}_j\|_2^2}{4D(t_q - t_k)}\right) \cdot \exp\left(-\lambda(t_q - t_k)\right) \tag{1}$$

for $q > k$. The formula can be simplified to

$$\mathrm{c}(\mathbf{p}_i, \mathbf{p}_j, t_q) = \frac{1}{(4\pi D(q-k)\Delta t)^{N/2}} \cdot \exp\left(-\frac{d_{i,j}^2}{4D(q-k)\Delta t}\right) \cdot \exp\left(-\lambda(q-k)\Delta t\right). \tag{2}$$

After the substitution $a = 4D\Delta t > 0$, $b = \lambda\Delta t > 0$ we obtain resulting activation formula

$$\mathrm{c}(\mathbf{p}_i, \mathbf{p}_j, t_q) = (\pi a(q-k))^{-N/2} \cdot \exp\left(-\frac{d_{i,j}^2}{a(q-k)} - b(q-k)\right). \tag{3}$$

When $\min(d_{i,j} \geq 1)$, then we suggest to use $a = 1, b = 1/10$ for the first experiments as will be demonstrated in next sections.

2.2 Complete Activation

The SOM learning is based on the substrate concentrations in q-th step in time t_q. This concentration represents the result of previous activation sequence $\varphi_1, \varphi_2, ...\varphi_{q-1}$ using single activation model (3). Due to linearity we can use the additivity principle and directly calculate the cumulative concentration in i-th neuron and step q

$$c_{i,q} = \sum_{k=1}^{q-1} c(\mathbf{p}_i, \mathbf{p}_{\varphi_k}, t_q - t_k) = \frac{1}{(\pi a)^{N/2}} \cdot \sum_{k=1}^{q-1} \frac{\exp\left(-\frac{d_{i,\varphi_k}^2}{a(q-k)} - b(q-k)\right)}{(q-k)^{N/2}}. \tag{4}$$

Resulting formula consists of all concentration information which are necessary for the SOM learning. Therefore the concentration $c_{i,q}$ is only a function of activation history, SOM topology and parameters a, b. But the history is result of learning which will be studied in next section.

The full concentration profile in $2D$ pudding after 99 random activation steps ($q = 100$) is depicted on the right part of figure 1.

2.3 Diffusive Learning of SOM

Novel learning algorithm is completely devoted to Kohonen learning rules [6] as follows. The weight of i-th neuron is changed in q-th step by rule

$$\mathbf{w}_i(q) = \mathbf{w}_i(q-1) + \alpha(q) \cdot c_{i,q} \cdot (\mathbf{x}_q - \mathbf{w}_i(q-1)) \tag{5}$$

for $i = 1, ..., H$, $\mathbf{x}_q \sim \mathrm{U}(\mathcal{S})$ is uniformly selected pattern from $\mathcal{S}$, $c_{i,q}$ is substrate concentration according to (4) and $\alpha(q) > 0$ is ageing function which is supposed to be non-increasing. The winner is also selected according to Kohonen rule [6] as

$$\varphi_q \in \underset{k=1,...,H}{\arg\min} \|\mathbf{x}_q - \mathbf{w}_k\|_2. \tag{6}$$

The main difference between the traditional SOM learning [7] and our approach [8] is in the application of diffusive equation which generates the concentration profile (4). The learning feedback is driven by winner index φ_q from (6) which is used in next step of concentration calculations (4).

As in traditional SOM learning we have to initialize the weights [7] and use appropriate ageing strategy. We recommend generate the initial weights from the multivariate Gaussian distribution as

$$\mathbf{w}_i(0) \sim \mathrm{N}(\mathrm{E}\mathbf{X}, \mathrm{var}\mathbf{X}/100) \tag{7}$$

for $i = 1, ..., H$. The ageing function $\alpha(q)$ can be constant in the first experiments, but satisfying $\alpha(q) \cdot c_{i,q} \leq 1$ to avoid learning instability.

3 Stock Market States

In our previous studies [9, 10] we investigated the states of country economics using annual change of macroeconomical indicators and traditional principal component analysis. In this study we focused on the states of individual stocks [11] and their changes which are easy to analyse and visualise using Pudding SOM algorithm. The analysis of single stock is based on price time series

$$\{a_k\}_{k=0}^{D-1} \tag{8}$$

in period of D days where $a_k \geq 0$ is the stock price in k-th day. First, we have to calculate daily yield as

$$z_k = \Delta \ln a_k = \ln(a_{k+1}/a_k) \tag{9}$$

The hypothesis behind the study is that the state reconstruction brings more information than the individual daily price or yield. The stock state is defined using slighting window of length $w \in N$ as vector

$$s_k = (z_k, ..., z_{k+w-1}) \in \mathbb{R}^w \tag{10}$$

To avoid the effect of dimensionality curse we will study the states for $w <= 30$ respecting the habit of 5 market days per week we will study the market states for $w = 5, 10, 15, 20, 25, 30$ in fixed hexagonal topology of 19 nodes SOM.

3.1 Case Study: Stock Market in Period 2000 - 2018

As input data we use daily prices of 10 main indices index SP500 (USA), index Dow Jones 30 (USA), index Nasdaq (USA), index Russell 2000 (USA), index DAX (Germany), index CAC40 (France), index BEL20 (Belgium), index EURONEXT100 (Europe), index NIKKEI (Japan) and Hang Seng index (Hong Kong). We have used data from March 2000 to March 2018. Therefore the SOM consists of 43070 states from 4312 days. The diffusive learning and traditional Kohonen approach have been compared for $w \in slozenezavorky 5, 10, 15, 20, 25, 30$ which corresponds to working week period. Selected results are demonstrated in figure 2 as number of states in SOM nodes. The results for $w >= 10$ are very similar and we decided to analyse individual stocks only for diffusive learning and $w = 5$. As seen the highest concentration are in the central nodes which represent the average behaviour. Meanwhile the extreme case in perimeter nodes are not so frequent in all cases.

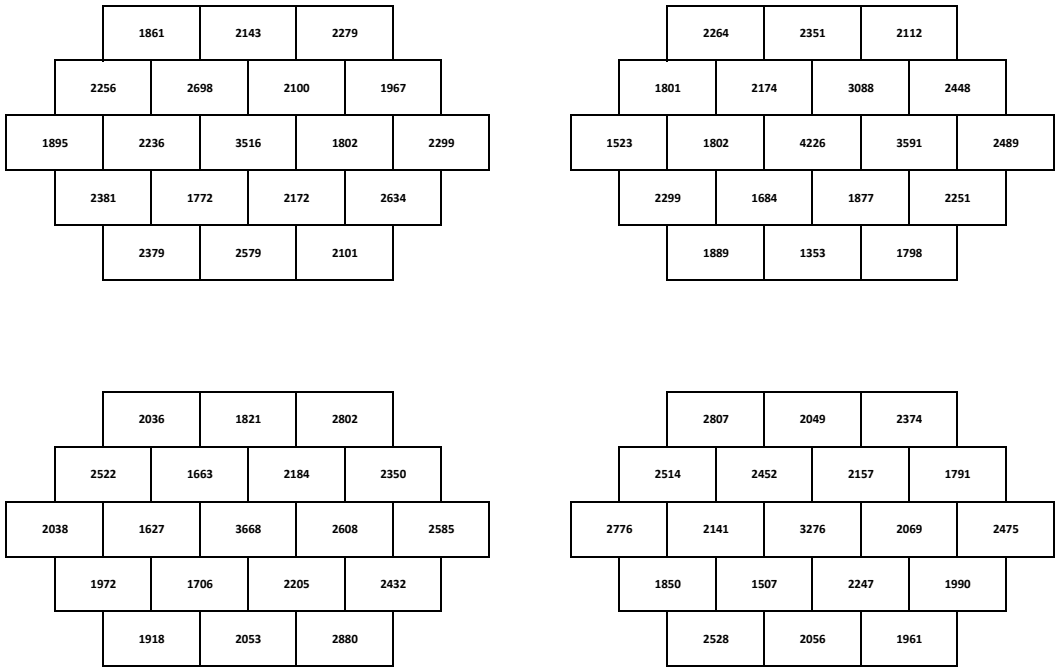


Figure 2: Resulting SOMs in case of diffusive learning with $w = 5$ (top left), $w = 10$ (top right) and Kohonen learning with $w = 5$ (bottom left), $w = 10$ (bottom right)

Using the result of learning we can trace the states of selected stock in selected years. The situation in the year 2008 i.e. the beginning of crisis is demonstrated in figure 3. The anomalous market behaviour caused higher frequency of extreme states in perimeter nodes as opposite to normal behaviour. The example of standard behaviour is captured in figure 4.

4 Conclusion

The new method of diffusive learning has been applied to stock market data and its results are similar to Kohonen SOM. The state reconstruction has been performed using one week history of individual stocks. There is significant difference between stocks states during crisis period and normal stock behaviour. The crisis can be characterized by high frequency of extreme states at the perimeter of SOM. So that the last year can be characterized by normal stock behaviour.

Acknowledgements

Supported by SGS17/196/OHK4/3T/14 grant of Czech Technical University in Prague.

References

[1] J. J. Thomson, On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure, Philosophical Magazine Series 6 7 (39) (1904) 237–265.

[2] A. Philippides, P. Husbands, M. O'Shea, Four-dimensional neuronal signaling by nitric oxide: a computational analysis, Journal of Neuroscience 20 (3) (2000) 1199–1207.

[3] D. Graupe, Deep Learning Neural Networks: Design and Case Studies, 2016.

[4] E. Oja, S. Kaski, Kohonen Maps, Elsevier Science, 1999.

[5] A. Rettberg, M. Zanella, M. Amann, M. Keckeisen, F. Rammig, Analysis, Architectures and Modelling of Embedded Systems: Third IFIP TC 10 International Embedded Systems Symposium, IESS 2009, Langenargen, Germany, September 14-16, 2009, Proceedings, IFIP Advances in Information and Communication Technology, Springer Berlin Heidelberg, 2009.

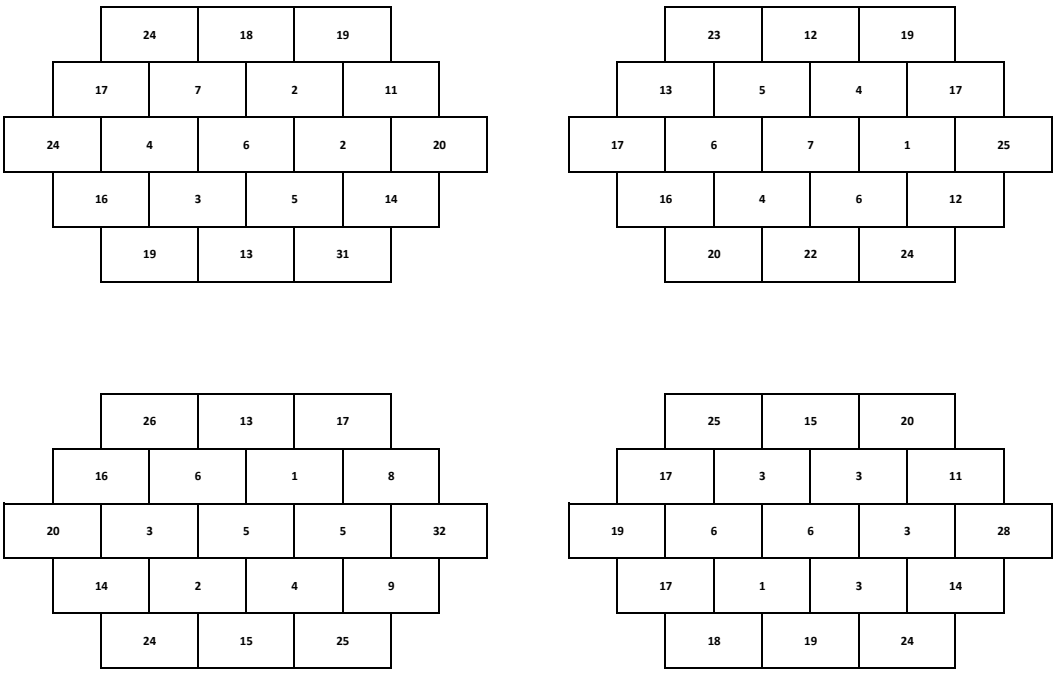


Figure 3: Resulting SOMs for SP 500 (left top), DAX (right top), NIKKEI (left bottom) and Hang Seng (right bottom) for diffusive learning with $w = 5$ in 2008

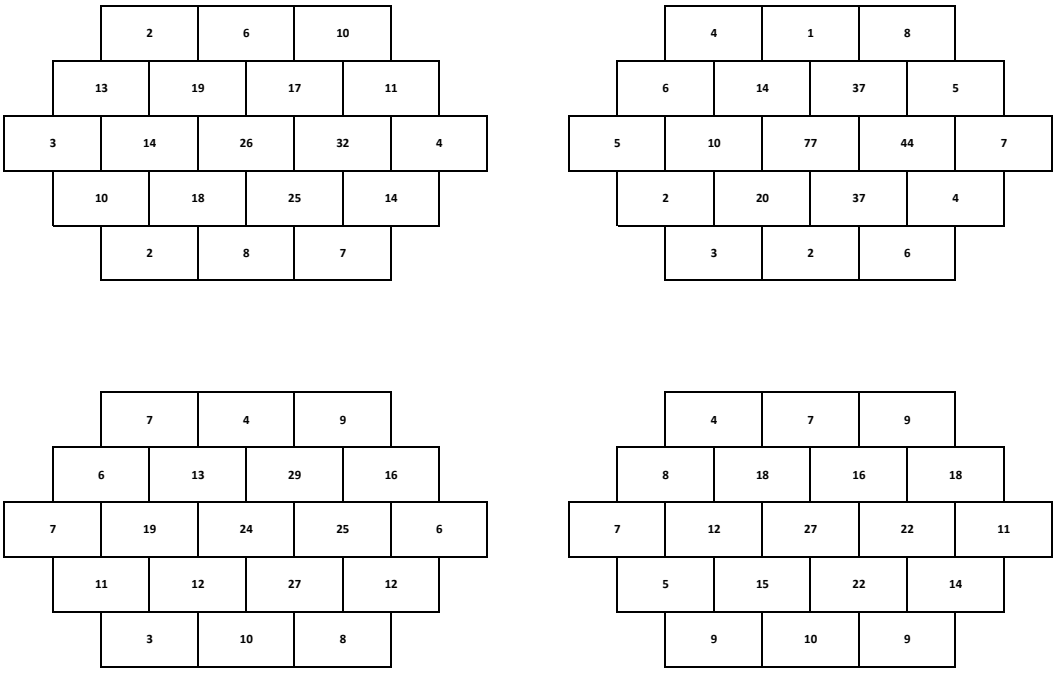


Figure 4: Resulting SOMs for SP 500 (left top), DAX (right top), NIKKEI (left bottom) and Hang Seng (right bottom) for diffusive learning with $w = 5$ in 2017

[6] T. Kohonen, Self-Organizing Maps, Springer Series in Information Sciences, Springer Berlin Heidelberg, 2012.

[7] E. Alonso, Computational Neuroscience for Advancing Artificial Intelligence: Models, Methods and Applications: Models, Methods and Applications, Premier reference source, Medical Information Science Reference, 2010.

[8] R. Hrebik, J. Kukal, Self-organization via diffusion modelling, IEEE Transactions on Neural Networks and Learning Systems (2018) submitted.

[9] R. Hrebik, J. Kukal, Multivarietal data whitening of main trends in economic development, in: Mathematical Methods in Economics, University of West Bohemia, Plze, 2015, pp. 279–284.

[10] R. Hrebik, J. Kukal, The economics and data whitening: Data visualisation, in: Federated Conference on Software Development and Object Technologies, Springer, 2015, pp. 91–101.

[11] M. Dlask, J. Kukal, Correlation dimension as a measure of stock market variability, in: Mathematical Methods in Economics, University of Hradec Kralove, Hradec Kralove, 2017, pp. 119–124.

Multiple criteria methods for definition of charging stations for freight transport

Michal Husinec[1], Tomáš Šubrt[2]

Abstract. Our paper deals with the technical definition of charging stations for electric vehicles. The scope for the possible use of freight electric vehicles in road transport is limited by the lack or absence of a charging station network. The analyzed problem is based on the limited infrastructure of the charging stations available to the trucks and the time needed to charge the battery. Nowadays existing charging stations for passenger cars cannot be used by trucks. The subject of the article is the application of a multi-criteria analysis method ANP for the purpose of locating charging stations, in terms of availability of energy network sources and efficiency of use during charging. The goal of the solution is also the efficient use of resources and the mitigation of environmental impacts in locations of new charging stations. The work provides a comprehensive overview of the use of modern methods and their benefits in their field of use.

Keywords: Multicriteria decision, analytic network process, charging stations, electric vehicle

JEL Classification: C44
AMS Classification: 90B50, 91B06

1 Introduction

In the field of sustainable road transport, the possible use of electric trucks as an alternative for conventional vehicles has been discussed in recent years. The area of possible use of vehicles is limited by the lack of or completely missing trucks charging stations infrastructure and by the overall time needed for charging the batteries.

The main reason for introducing electromobility is its minimal environmental impact [3]. Studies on the use of the energy network indicate the benefits in the area of charging at night, when the electric vehicles are mainly recharged [11]. Overnight charging helps to balance the power supply in the energy network by consuming surpluses from production. Together with the development of electric vehicles, we should work with the idea of a smart energy network that will ensure energy for these vehicles [8] in the future. Improving the efficiency of car is linked to the network efficiency increasing, thereby achieving significantly lower energy consumption and reducing carbon emissions produced by the energy production. The main constraints are the vehicle's range of batteries. Manufacturers indicate that for vehicles with a total weight of 3.5t and above is the range around 120 km [7] which may vary in relation to the laden capacity and climatic conditions. Based on the long-term tests of Mercedes-Benz, the range in urban areas is sufficient [6]. Nevertheless, what happens when the vehicle has to reach outside this radius and we need it to return to the central depot. Or, as a result of congested infrastructure capacity, the vehicle will get stuck in a traffic jam. The vehicle must then locate the charging station and recharge the battery. The company FUSO, a subsidiary of Mercedes-Benz, with its product called E-Canter indicates the battery charge up to 80% in one hour if the direct current is used for charging [6]. Ideal vehicle recharging is a type 2 (EU compatible) CCS connector. This connector allows you to charge the battery with up to 200A [1]. Currently there are 43 charging stations in the Czech Republic with the possibility of connecting this connector. Unfortunately, these stations are only usable for recharging passenger cars due to their location in city centers. Overall, the network of charging stations for trucks is inadequate, therefore there is a room for system analysis of the definition of the charging station and the optimization of the charging stations location. The location should include criteria for assessing the ideal location. One of the evaluation criteria should also be the location near the logistic parks. In this context, hybrid algorithms based on genetic algorithms and the conventional PSO (Particle Swarm Optimization) algorithm [2] appear for optimization. The suggested GAIPSO method increased the functionality of algorithm PSO and shows a new optimal charging station locations strategy. The designed algorithm fulfilled the required criteria for the minimum number of charging stations and the optimal energy profile in the energy network. In terms of definition, the charging stations of the cargo-related electric

[1] Czech University of Life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Prague, Czech Republic, e-mail: husinecm@pef.czu.cz
[2] Czech University of Life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Prague, Czech Republic, e-mail: subrt@pef.czu.czUniversity/Institute, department, address, e-mail.

vehicles could be integrated to the category of logistical objects with the possibility of using an algorithm to localize objects with charging stations parameters restrictions. The placement issue depends on the complexity of the logistic network which will be multi-layered (indirect relationship: Suppliers-Intermediate points-recipients). The problem with locations of warehouses is a multi-objective optimization problem, which depends on quantitative and qualitative criteria [9]. The number of criteria emphasizes the complexity of the problem resulting from the possible location of charging stations for which multicriterial decision-making methods can be used in order to select a solution

Multiobjective decision problem for the definition of the charging station represents the analysis selected by the method ANP (Analytic Network process) and it is very difficult to make a decision on the correct definition of the charging station without using the ANP method (the specification of the criteria and Sub-criteria are not mutually independent), which will be suitable for the availability of energy network resources and the efficiency of use during charging. For the ANP method, the importance of all including elements is determined by pair comparison on a scale of 1-9, similar to the AHP method. The difference lies in the elements clusters interdependence rating. From methodological perspective we present an approach based on the ANP method for evaluation of electric charging stations. The essence of using the ANP model is the efficient use of resources and mitigation of environmental impacts in the construction of new charging.

2 ANP as a tool for evaluating the criteria of charging stations

Models of Multicriterial decision-making are now used in many areas for quantification, comparison and evaluation. The Analytical Network process (ANP) is one of the most effective tools in cases where the interaction between qualitative and quantitative factors creates a network structure [13], [14]. It dissolves the decision-making problem into a network of sub-problems, which are then analyzed and evaluated. Network structuring of the decision-making problem helps to express situations in which decision-making criteria are grouped by sub-criteria for interaction models. Criteria rectangles can also represent individual clusters (clusters), e.g. criteria of the decision-making problem. The method of analysis of network process is the method newly derived by T. L. Saaty in 1996 [12]. The method was primarily derived from the AHP method, but in a fact, AHP is a more versatile method when ANP is just one of their specific instances. The main benefit of the method is that it allows to model and evaluate interdependence (influencing) between alternatives, criteria and sub-criteria, or sub-criteria of each other. The general model of ANP has thus four-level hierarchical structure,

1. Objective of the analysis or decision
2. Criteria Group
3. Sub criteria Group, where the sub-criteria tend to include barriers to achieving goals.
4. Group of alternatives.

In terms of the quality of feedback, the ANP models can be divided into two groups: first the FSM (Feedback System models) with the feedback within the sub-criteria group, and the second the SSM (Series System models), without these feedbacks. Essential for determining the definition of the charging station are the feedbacks of clusters, which are connected to each other (one-to-one) into a large complex network model. We assume that individual criteria interact as well as the criteria interact with each other. This means that the hierarchical structure could be transformed into a network structure and the ANP model seems to be a very suitable tool for resolving this problem. ANP Super-matrix (not weighted, weighted, marginal) with possible grouping of interactions and influences shall be defined and calculated and the most appropriate project design shall be selected by synthesis (adding) of all the control criteria. The calculation itself has been done by the software Superdecision [15].

$$\mathbf{W} = \begin{array}{c} \\ C \\ S \\ P \end{array} \begin{array}{c} \begin{array}{ccc} C & S & P \end{array} \\ \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{W_C} \\ \mathbf{W_S} & \mathbf{W_{S^*}} & \mathbf{0} \\ \mathbf{0} & \mathbf{W_A} & \mathbf{0} \end{pmatrix} \end{array}$$

In which:

- WC the balance criteria matrix with respect to the definition of the charging station
- WS is the sub-criteria balance matrix with respect to the criteria
- WS* is the weight matrix of sub-criteria with respect to each other
- WA is the balance matrix solution with respect to sub-criteria

Level of criteria in the ANP process is designed by SMART principle. The aim is a general declaration of the desired outcome with one or more sub targets that define exactly what the resulting charging station should look like. Criteria and sub-criteria that are taken into account to select the location of the charging stations have been

derived from Wu et al [16] and customized for freight transport in Czech conditions. Thus some criteria were refined and expanded (Geographical location) and some of them were modified and enhanced (original Wu text was designed for Chinese conditions). Among most modified ones Energy saving (stability) can be named. Here the possibility of creating an autonomous station independent of the energy network was taken into account. All the criteria are complex and are composed as a mix of demand based (e.g. Convenience of transportation, Location..) supply based (e.g. Rechargeable Capability, Capacity expansion...) and general ones (Environmental impacts, Support from local government ..)

1. **Economic factors**

- Construction costs – costs associated with the purchase of land, project documentation, construction of the charging station. Measurable in Euro;
- Operating costs - costs including all operating charges related to daily operation. Operating costs are important in terms of financial gain. Measurable in Euro;
- Return on investment (ROI) – relevant in terms of cost and operating income assessment. The most important economic criterion in the commercial sphere. Measurable in months;

2. **Technical factors**

- Distance from the distribution station – location near the distribution station affects loss of power transmission. Measurable in km;
- Impact on the energy network – relevant for the safe operation of the energy network. Measurable in MW;

3. **Service availability**

- Rechargeable Capability-Maximum number of charging stations in a single moment. This sub-criterion is related to the number of vehicles that can be charged. Measurable in number of charging stations.
- Availability (Convenience of transportation) – criterion important from the point of view of the available road network by a lorry. Measurable in the number of accessible roads;

4. **Social factors**

- Capacity expansion – a necessary requirement in terms of further increases of freight electric vehicles. Measurable in volume of newly build charging stations;
- Local position – relevant requirement of impact on the potential health problems of the population (electromagnetic fields, traffic growth). Measurable in the number of permits issued for the construction;
- Support from local government – this attribute may include possible grant titles for construction (National Action Plan for Clean Mobility). Measurable in the number of issues released for construction;

5. **Environmental impacts**

- Environmental impact – the necessary environmental impact assessment (EIA process). Measurable in carbon footprint CO2;
- Energy saving (stability) – the possibility of creating an autonomous station independent of the energy network or creating a hybrid model that could respond to excess of energy in the network and thus compensate for fluctuations. Measurable in MW.
- Impact on the energy network – an immediate demand for performance (charging) can destabilize the energy network. Measurable in MW;

6. **Location (geographic location)**

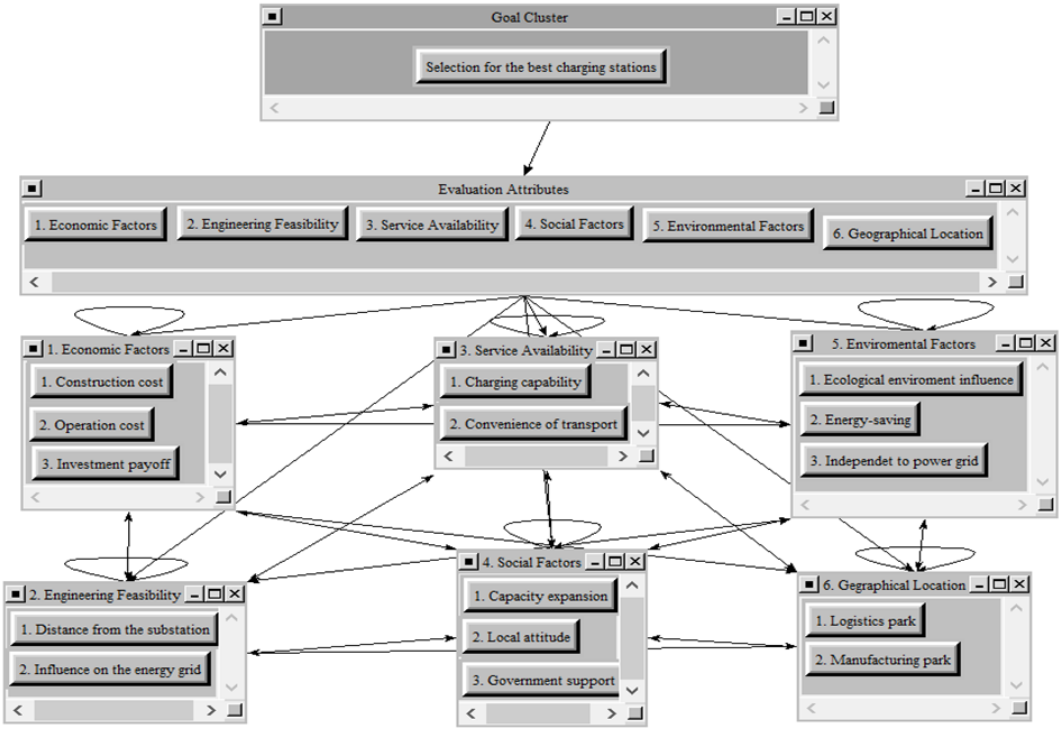
- Logistic Parks – relevant from the point of view of logistics chains engagement accepting places with a high concentration of vehicles used for products (goods) transportation. Measurable in the number of available logistic parks;
- Logistic objects – places with concentration of production plants. Measurable in the number of production plants and parking areas with logistic services.

3 ANP Model for Charging Station Definition

The ANP software tool Superdecision was used to build a network dependency between criteria and sub-criteria and to calculate necessary weights. The first level of the hierarchical network under the Goal node represents the level of criteria as unique clusters (there are no Mutual Dependence). The balance of the criteria was set identically to 0.167.

Other level is represented by individuals criteria (clusters) with sub-criteria that are a key part of the process of evaluating the importance of each sub-criterion within the ANP process. These relations are based on the experience of other authors or subjective assessment of the author in terms of importance (example of mutual influence – the distance criterion from the distribution station will affect the construction costs. However, it does not have a direct effect on the station availability criterion by a lorry. Vehicle availability, construction costs, operating costs and the capacity of the charging stations affect the criteria Return of Investment). Most important criteria should be economic factors, technical feasibility in terms of the energy network, or possible construction of an autonomous station, the number of charging stations. This criteria will evaluate the overall return on

investment. Relations within the clusters are most current within economic, technical, and location situations for location evaluation for possible location of charging stations to already established logistic parks or other industrial premises (manufacturing area, logistics terminal and places from high concentration freight transport). Achieving certain sub-criteria precludes the achievement of others or, conversely to achieve one sub-criteria accelerates the achievement of another sub-criteria. If there are any sub-criterion relevant, it is important to make it realistic. Similar relations were observed within other clusters (Picture 1). The current weights are based on an expert evaluation and are calculated by using of a Saaty matrix that is an integral part of the software Superdecision [15].



Picture 1 Network model of Charging Station Definition (software Superdecision)

3.1 ANP model Results

The first ANP result, un-weighted super-matrix for equal criteria weights, gives a good impression of clusters, established connections and their evaluation by weights.

		Economic Factors	Engineering Feasibility	Service Availability	Social Factors	Enviromental Factors	Gegraphical Location
Economic Factors	**0,167**	0,00	0,00	0,00	0,00	0,00	0,00
Engineering Feasibility	**0,167**	0,00	0,00	0,00	0,00	0,00	0,00
Service Availability	**0,167**	0,00	0,00	0,00	0,00	0,00	0,00
Social Factors	**0,167**	0,00	0,00	0,00	0,00	0,00	0,00
Enviromental Factors	**0,167**	0,00	0,00	0,00	0,00	0,00	0,00
Gegraphical Location	**0,167**	0,00	0,00	0,00	0,00	0,00	0,00
Contruction cost	0,00	**0,57**	0,00	0,00	0,00	0,00	0,00
Operation cost	0,00	**0,29**	0,00	0,00	0,00	0,00	0,00
Investment payoff	0,00	**0,14**	0,00	0,00	0,00	0,00	0,00
Distance from the substation	0,00	0,00	**0,34**	0,00	0,00	0,00	0,00
Influence on the energy grid	0,00	0,00	**0,66**	0,00	0,00	0,00	0,00
Charging capability	0,00	0,00	0,00	**0,66**	0,00	0,00	0,00
Convenience of transport	0,00	0,00	0,00	**0,34**	0,00	0,00	0,00
Capacity expansion	0,00	0,00	0,00	0,00	**0,58**	0,00	0,00
Local attitude	0,00	0,00	0,00	0,00	**0,24**	0,00	0,00
Government support	0,00	0,00	0,00	0,00	**0,18**	0,00	0,00
Ecological enviroment influence	0,00	0,00	0,00	0,00	0,00	**0,41**	0,00
Energy-saving	0,00	0,00	0,00	0,00	0,00	**0,33**	0,00
Independet to power grid	0,00	0,00	0,00	0,00	0,00	**0,26**	0,00
Logistics park	0,00	0,00	0,00	0,00	0,00	0,00	**0,50**
Manufacturing park	0,00	0,00	0,00	0,00	0,00	0,00	**0,50**

Table 1 First part of unweighted matrix – goal node and criteria

The ANP results also deals with final weights calculations using limit matrix. Limit calculation brings following weights for SMART sub-criteria (Table 2.). As supposed, most important sub-criteria are those, usually mentioned ahead within Service availability, Engineering feasibility and Economic factors.

Sub-criteria	Weight
1. Charging capability	0,179
1. Distance from the substation	0,115
1. Capacity expansion	0,096
1. Logistics park	0,093
1. Construction cost	0,086
2. Operation cost	0,066
2. Influence on the energy grid	0,061
2. Energy-saving	0,061
2. Local attitude	0,060
3. Government support	0,054
2. Manufacturing park	0,039
1. Ecological enviroment influence	0,034
3. Independet to power grid	0,024
2. Convenience of transport	0,020
3. Investment payoff	0,013

Table 2 Sub criteria limit weights

Figure 1 shows limit scales sub-criteria in a graphic form. We applied the weighting limits to the existing logistic areas to find a suitable location for the charging station. The solution to the possible location is still distorted by the capacity of the energy network, for which it is necessary to handle more detailed impact of construction and immediate unit sampling when charging stations are connected. It is surprising that our originally preferred criteria and sub-criteria have a lower marginal weight than expected. The ANP method uncovered other major sub-criteria that are essential for the correct charging station definition.

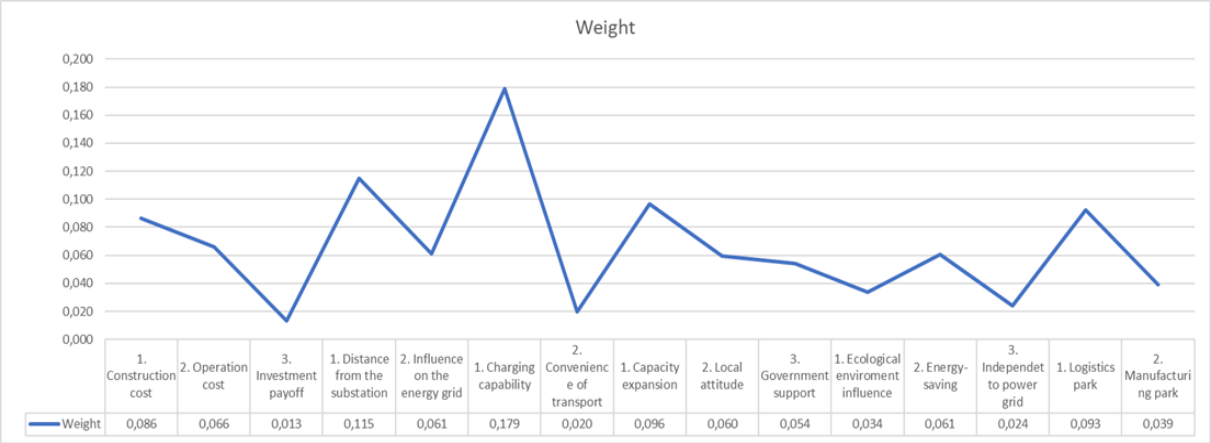


Figure 1 Sub-criteria limit weights - chart view

4 Conclusion

The aim of the paper was to analyze the quality and quantity of the various criteria which are essential to define the charging station for electric vehicles. The problems of freight electro-vehicles charging stations unlike the charging stations for passenger cars have not yet been analyzed in detail and thus ANP analysis with a SMART approach to evaluating the criteria, here can be beneficial.

1. The methodology used appears to be suitable for the definition of the charging station according to different criteria.
2. The ANP method allows description and research of complex dependencies between important criteria and sub-criteria from different viewpoints. Network dependencies are typical for this issue.

Our future research will focus on the weight of the criteria and the assessment of the actual design of the charging station definition. These weights must be estimated according to the expert's judgement, therefore a set of criteria requires a soft system approach. Further relevant criteria for the definition of the charging station can be added. In view of the future fast-growing support for electromobility by local authorities, the aid criterion is likely to be shifted from a cluster of social factors into economic factors. Other new criteria can be e.g.: the increase in working opportunity in the region and the development of opportunities for small enterprises offering additional services (accommodation, gastronomy, etc.).

Acknowledgements

The contribution is also supported by the grant Project IGA PEF, Czech University of Agriculture in Prague, Czech Republic. No. 20181020 " Využití vícekriteriálních metod pro definici nabíjecích stanic nákladních elektromobilů" ("Multiple criteria methods for charging stations definition used for freight transport").

References

[1] Association for electromobility of the Czech Republic, [Online], Available: http://www.elektromobily-os.cz/úvod-1 (07.01.2018)
[2] Awasthi, A. et al. (2017) „Optimal planning of electric vehicle charging station at the distribution system using hybrid optimization algorithm", Energy, 133, s. 70–78. doi: 10.1016/j.energy.2017.05.094.
[3] Behnke, M. a Kirschstein, T. (2017) „The impact of path selection on GHG emissions in city logistics", Transportation Research Part E: Logistics and Transportation Review. Pergamon, 106, s. 320–336. doi: 10.1016/J.TRE.2017.08.011.
[4] Boysen, N., Briskorn, D. a Emde, S. (2018) „Scheduling electric vehicles and locating charging stations on a path", Journal of Scheduling. Springer US, 21(1), s. 111–126. doi: 10.1007/s10951-017-0538-9.
[5] Deb, S. et al. (2018) „Impact of Electric Vehicle Charging Station Load on Distribution Network", Energies. Multidisciplinary Digital Publishing Institute, 11(2), s. 178. doi: 10.3390/en11010178.
[6] E-Canter, [Online], Available: http://www.ecanter.cz/otazky-a-odpovedi (07.01.2018)
[7] Electric mobility - Daimler Global Media Site, [Online], Available: http://media.daimler.com/marsMediaSite/en/instance/ko/Electric-mobility.xhtml?oid=9265809 (07.01.2018)
[8] Huang, K., Kanaroglou, P. a Zhang, X. (2016) „The design of electric vehicle charging network", Transportation Research Part D: Transport and Environment. Pergamon, 49, s. 1–17. doi: 10.1016/J.TRD.2016.08.028.
[9] Jacyna-Gołda, I. a Izdebski, M. (2017) „ScienceDirect The Multi-criteria Decision Support in Choosing the Efficient Location of Warehouses in the Logistic Network", Procedia Engineering, 187, s. 635–640. doi: 10.1016/j.proeng.2017.04.424.
[10] Lee, J. a Madanat, S. (2017) „Optimal design of electric vehicle public charging system in an urban network for Greenhouse Gas Emission and cost minimization". doi: 10.1016/j.trc.2017.10.008.
[11] Montoya, A. et al. (2017) „The electric vehicle routing problem with nonlinear charging function", Transportation Research Part B, 103, s. 87–110. doi: 10.1016/j.trb.2017.02.004.
[12] Saaty, T. L. (1994) „How to Make a Decision: The Analytic Hierarchy Process", Interfaces, 24(6), s. 19–43. doi: 10.1287/inte.24.6.19.
[13] Saaty, T. L. a Hu, G. (1998) „Ranking by Eigenvector versus other methods in the Analytic Hierarchy Process", Applied Mathematics Letters, 11(4), s. 121–125. doi: 10.1016/S0893-9659(98)00068-8.
[14] Saaty, T. L. (2008) „Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process", Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas, 102(2), s. 251–318. doi: 10.1007/BF03191825.
[15] SuperDecisions - Software for decision-making, (21.03.2018) http://www.superdecisions.com/.
[16] Wu, Y. et al. (2016) „Optimal Site Selection of Electric Vehicle Charging Stations Based on a Cloud Model and the PROMETHEE Method", Energies. Multidisciplinary Digital Publishing Institute, 9(3), s. 157. doi: 10.3390/en9030157.

What drives the estimation results of DSGE models? Effect of the input data on parameter estimates

Jakub Chalmovianský[1]

Abstract. In this contribution, I compare three different Bayesian dynamic stochastic general equilibrium (DSGE) models in a simulation-estimation exercise. This exercise is aimed at revealing the capabilities of these models to re-estimate, during the estimation phase, values of parameters previously set in the simulation phase. The first model is the renowned work of Smets and Wouters [7]. The second one is the rather small DSGE model of a closed economy with search and matching frictions on labour market proposed by Lubik [5]. The third one is based on the paper written by Sheen and Wang [6], where they introduce a model of a small open economy with various labour market frictions.
The aim of this contribution is to examine how the complexity of the model and the amount of information needed, represented by the number of observations in the observables, affect the results when the parameters are estimated. At first, I shortly introduce all presented models. Based on the given calibration, trajectories of main endogenous variables are simulated. These simulated trajectories with a various number of observations are then used as observables for estimation of the model parameters to reveal how rich information is needed for each model to properly identify its parameters.

Keywords: Bayesian estimation, DSGE model, Matching moments, Model simulation, Parameter identification.

JEL classification: C32, C52
AMS classification: 91B51

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have become a standard tool (not only) for monetary policy analysis of aggregate economic phenomena or investigation of the mutual interaction between monetary and fiscal policy. Thanks to the increased computational power and enhanced numerical approximation methods, estimation of rather complex DSGE models is feasible. Nevertheless, their complex structure and increasing number of parameters together with only a limited number of suitable observable macroeconomic time-series or length of these time-series available give rise to the question of identifiability of their parameters during the estimation. This issue is of special interest in the region of Central Europe where the number of observations for macroeconomic indicators is often limited. For more details on the identification issues in DSGE models, I refer interested readers to the work of Canova and Sala [1], the contribution of Iskrev [3], or the paper by Koop, Pesaran and Smith [4].

In this contribution, I investigate the stability of three different Bayesian DSGE models in a simple simulation–estimation exercise. I focus on the capabilities of the selected DSGE models to re-estimate their parameters during the estimation phase. First of the models is the renowned medium-scale model of Smets and Wouters [7]. The second one is a rather small DSGE model of a closed economy with search and matching frictions on labour market proposed by Lubik [5]. The third model is based on the paper written by Sheen and Wang [6], where authors, in the end, present four different specifications of their model of a small open economy with combinations of the following two aspects: a model with/without hiring costs, with real/nominal wage rigidities.

The simulation–estimation exercise consists of two stages. During the first stage, based on the given calibration, trajectories of main endogenous variables are simulated. These simulated trajectories are in the second stage used as observables for estimation of the model parameters. Three different scenarios for the length of observable time series are used for each model (100, 75, and 50 observations respectively) to reveal how the extent of the given information may affect the ability of each model to properly identify its parameters. For discussion about the effect of the selection of observables on DSGE estimation results, I refer interested readers to see the paper written by Guerrón–Quintana [2].

The aim of this contribution is to examine how the complexity of a model and the amount of information needed, represented by the number of observations in the input observables, affect the resulting parameter esti-

[1]Masaryk University, Faculty of Economics and Administration, Lipová 41a, Brno, Czech Republic, chalmovianskyy@mail.muni.cz

mates. At first, I briefly introduce all presented models. Subsequently, the summary of calibrated parameters and estimation methods used are presented, followed by the discussion over results of the proposed simulation–estimation exercise for my models of interest. Summary of main findings concludes.

2 Evaluated models

2.1 Model Smets, Wouters (2007)

The model proposed by Smets and Wouters in [7] is an example of estimated New Keynesian DSGE model of medium-scale. It is one of the most cited DSGE models in recent years and shortly after its publication it has become a work-horse for monetary policy analysis. This model was initially developed and used in European Central Bank, however, the referenced paper presents results from the estimation of the US economy. It allows for both price and wage rigidities, habit formation in consumption, and investment adjustment cost. There are seven different structural shocks: total factor productivity shock, ε_t^a, risk-premium shock, ε_t^b, investment-specific technology shock, ε_t^i, wage mark-up shock, ε_t^w, price mark-up shock, ε_t^p, exogenous spending shock, ε_t^g, and monetary policy shock, ε_t^r. Overview of main equations of the linearized model is as follows:

$$y_t = \phi_p \left(\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^a\right) \tag{1}$$

$$k_t^s = k_{t-1} + z_t \tag{2}$$

$$z_t = z_1 r_t^k \tag{3}$$

$$r_t^k = -\left(k_t - l_t\right) + w_t \tag{4}$$

$$y_t = c_y c_t + i_y i_t + z_y z_t + \varepsilon_t^g \tag{5}$$

$$c_t = c_1 c_{t-1} + (1-c_1) E_t c_{t+1} + c_2 (I_t - E_t I_{t+1}) - c_3 (r_t - E_t \pi_{t+1} + \varepsilon_t^b) \tag{6}$$

$$i_t = i_1 i_{t-1} + (1-i_1) E_t i_{t+1} + i_2 q_t + \varepsilon_t^i \tag{7}$$

$$q_t = q_1 E_t q_{t+1} + (1-q_1) r_{t+1}^k - (r_t - E_t \pi_{t+1} + \varepsilon_t^b) \tag{8}$$

$$k_t = k_1 k_{t-1} + (1-k_1) i_t + k_2 \varepsilon_t^i) \tag{9}$$

$$\mu_t^p = \alpha (k_t - l_t) + \varepsilon_t^a - w_t \tag{10}$$

$$\pi_t = \pi_1 \pi_{t-1} + \pi_2 E_t \pi_{t+1} - \pi_3 \mu_t^p + \varepsilon_t^p \tag{11}$$

$$\mu_t^w = w_t - mrs_t = w_t - \left(\sigma l_t - \frac{1}{1-\lambda/\gamma}(c_t - \lambda c_{t-1})\right) \tag{12}$$

$$w_t = w_1 w_{t-1} + (1-w_1) E_t (w_{t+1} + \pi_{t+1}) - w_2 \pi_t + w_3 \pi_{t-1} - w_t \mu_t^w + \varepsilon_t^w \tag{13}$$

$$r_t = \rho r_{t-1} + (1-\rho)(r_\pi \pi_t + r_y (y_t - y_t^p)) + r_{\Delta y} \left[(y_t - y_t^p) - (y_{t-1} - y_{t-1}^p)\right] + \varepsilon_t^r \tag{14}$$

$$\varepsilon_t^a = \rho^a \varepsilon_{t-1}^a + \eta_t^a \tag{15}$$

$$\varepsilon_t^g = \rho^g \varepsilon_{t-1}^g + \eta_t^g + \rho^{ga} \eta_t^a \tag{16}$$

$$\varepsilon_t^b = \rho^b \varepsilon_{t-1}^b + \eta_t^b \tag{17}$$

$$\varepsilon_t^p = \rho^p \varepsilon_{t-1}^p + \eta_t^p - \mu_p \eta_{t-1}^p \tag{18}$$

$$\varepsilon_t^w = \rho^w \varepsilon_{t-1}^w + \eta_t^w - \mu_w \eta_{t-1}^w \tag{19}$$

$$\varepsilon_t^r = \rho^r \varepsilon_{t-1}^r + \eta_t^r \tag{20}$$

where y_t is GDP, l_t is labour input, k_t^s is capital in use which follows the evolution described in (2) with z_t denoting capacity utilisation and r_t^k as marginal productivity of capital, w_t describes real wage, c_t is consumption, i_t investment, E_t is expectations operator, r_t is the interest rate on a one-period safe bond, π_t denotes inflation rate, q_t is the real value of the existing capital stock, installed capital is denoted by k_t, μ_t^p is mark-up price over marginal cost, similarly, μ_t^w denotes wage mark-up, where mrs_t is marginal rate of substitution between working and consuming, and finally y_t^p is the potential level of output. The exogenous disturbances are then assumed to be orthogonal and described as autoregressive processes of the first order, see eq. (15) - (20). For detailed derivation of the model equations please refer to [7].

2.2 Model Lubik (2009)

The second model is the least complex one. It is the model proposed by Lubik in [5] and it represents a small-scale DSGE model that incorporates search and matching frictions on labor market and right-to-manage wage bargaining process between firms and households. Thanks to the search and matching process, creation of seeker-vacancy pair between job searchers and firms looking for workers is time-consuming and costly. As the model, I applied, is identical to the original one, I will present here only a brief overview of its main equations. For details on the model structure, estimation techniques etc. please refer to [5].

$$\lambda_t = C_t^{-\sigma} \tag{21}$$

$$m\left(u_t, v_t\right) = \mu_t u_t^{\xi} v_t^{1-\xi} \tag{22}$$

$$q\left(\theta_t\right) = m\left(u_t, v_t\right)/v_t \tag{23}$$

$$\theta_t = \frac{v_t}{u_t} \tag{24}$$

$$n_t = 1 - u_t \tag{25}$$

$$n_t = (1-\rho)\left(n_{t-1} - v_{t-1} q\left(\theta_{t-1}\right)\right) \tag{26}$$

$$y_t = A_t n_t^{\alpha} \tag{27}$$

$$y_t = \left(\frac{p_t}{P_t}\right)^{-1-\varepsilon_t} Y_t \tag{28}$$

$$\tau_t = \alpha \frac{y_t \varepsilon_t}{n_t(1+\varepsilon_t)} p_t - w_t + (1-\rho) E_t \beta_{t+1} \tau_{t+1} \tag{29}$$

$$\kappa v_t^{\psi-1} = (1-\rho) q_t E_t \beta_{t+1} \tau_{t+1} \tag{30}$$

$$\beta_t = \beta \lambda_t / \lambda_{t-1} \tag{31}$$

$$w_t = \eta \left(\alpha \frac{y_t \varepsilon_t}{n_t(1+\varepsilon_t)} p_t + \kappa v_t^{\psi-1} \theta_t\right)(1-\eta)\left(b + \chi_t C_t^{\sigma}\right) \tag{32}$$

$$Y_t = C_t + \frac{\kappa}{\psi} v_t^{\psi} \tag{33}$$

The first-order condition for households is shown in (21), where λ_t is the Lagrange multiplier on the representative household's budget constraint, C_t is aggregate consumption, $\sigma \geq 0$ is the coefficient of relative risk aversion in the household's utility function. In equation (22), $m(u_t, v_t)$ represents the number of matches of unemployed job-seekers, u_t, and vacancies, v_t, matching shock μ_t measures the efficiency of the matching process, parameter $0 < \xi < 1$ stands for the match elasticity of unemployed. Further, $q\left(\theta_t\right)$ represents aggregate probability of filling a vacancy, with θ_t indicating labour market tightness, n_t is the employment rate, which develops over time as given by (26), with separation rate '$0 < \rho < 1$. For simplicity, this model assumes monopolistic behaviour of firms, with production (and demand) y_t described by production function (27), with stochastic process A_t describing an aggregate technology shock and $0 < \alpha \leq 1$, introducing curvature in production. Demand function of a representative firm is given by (28), with aggregate output, Y_t, price p_t set by firm, and aggregate price index P_t; ε_t stands for demand elasticity assumed to be an exogenous process. The first-order conditions of the firms are (29) and (30), where τ_t is the Lagrange multiplier associated with firms employment constraint, w_t denotes bargained wage, β_t is stochastic discount factor, E_t denotes the expectations operator, parameter $\kappa > 0$ is a scaling factor on vacancy creation, and $\psi > 0$ stands for elasticity of vacancy creation. Parameters $\eta \in [0, 1]$ and b in the expression for bargained wage (32) denote bargaining power of worker and unemployment benefits, financed by lump-sum tax T_t, respectively. The model is complemented by specifying properties of exogenous shocks. For logarithms of technology shock, A_t, labour shock, χ_t, demand shock, ε_t, and matching shock, μ_t, the independent autoregressive processes of the first order with residuals $\epsilon_t^i \sim N(0, \sigma_i^2)$ and coefficients ρ_i, where $i \in (A, \chi, \varepsilon, \mu)$ are defined.

2.3 Model Sheen, Wang (2014)

Third and the last studied model is the medium-scale model proposed by Sheen and Wang in [6], where the authors compared and assessed small open economy model with different labour market frictions. This model links the previous two, as it contains sticky prices and wages, habit formation in consumption and costly investment adjustment as the model by Smets and Wouters, together with the search and matching process on the labour market as in Lubik's model.

Authors, in fact, present the following four different specifications of one model:

- model with real wage rigidities and hiring costs (RWHC),
- model with real wage rigidities, without hiring costs (RWNOHC),
- model with nominal wage rigidities and hiring costs (NWHC),
- model with nominal wage rigidities, without hiring costs (NWNOHC),

In their original contribution, Sheen and Wang found out that model with real wage rigidities and hiring costs suits the Australian economy the best. However, in my contribution, I assess all four specifications. Because of the complexity of the model (linearized economy consists of more than 40 equations) no overview of the model is given at this point. I refer interested readers to find all information about the model in the original paper (see [6]).

3 Empirical results

Simulation–estimation exercise used in this contribution can be described by the following algorithm:

1. Calibrate the model parameters with the given values.
2. Take a random number generator and simulate new data, i.e. shocks and variable paths, with the number of observations equal to $Q = Q_0 + Q_1$, where Q_1 is equal to the length of interest (100, 75, or 50 observations), using the values of parameters from the previous step.
3. Take the smoothed paths of endogenous variables, discard the first Q_0 generated values to avoid the potential distortion by initial values.
4. Use selected endogenous variables from the previous step as observables, and estimate the model with these trajectories used for input.
5. Repeat steps 2, 3 and 4 M times.
6. Compute the mean, and other moments from your estimated trajectories and compare them with the values from simulated trajectories.

The length of data sample used in this exercise is equal to $Q_1 \in \{100, 75, 50\}$ and I set $Q_0 = 100$ to minimize the impact of initial values. The number of iterations is set to $M = 20$, except the Sheen and Wang model, where this would be enormously time-consuming, therefore only 1 iteration was produced for each specification of this model. Table 1 summarizes parameter settings used for the simulation stage. In all models, parameter settings used are as close to the values originally proposed by authors as possible.

In Table 2, I compare means and standard deviations of trajectories, which authors originally used as observable input, simulated by the model with the moments in smoothed trajectories after estimation. These moments are based on 20 repetitions of the simulation–estimation exercise, for the model by Sheen and Wang only one realization of the trajectories of interest is available.

As can be seen from the figures in Table 2 number of observations does not have any major impact on the consistency of moments from estimated trajectories with their simulation counterparts. Figures are naturally different when 100, 75, or 50 observations are taken into account, however, in the vast majority of cases, moments from estimated trajectories follow the respective moments from simulated trajectories.

Focusing on the complexity of the model and its effect on the ability of the model to replicate its simulated data, no definitive answer can be given. Nevertheless, an intuitive expectation of more dissimilarities in more complex models does not seem to be applicable in this case. The best performance from this point of view shows the Smets and Wouters model, which is a medium-scale model. The small-scale model of Lubik seems to have a bit more difficulties with reproducing the simulated trajectories during the estimation. However, differences are still rather reasonably small. The biggest non-uniformity is represented by the most complex Sheen and Wang model. However, results from only one replication of the simulation–estimation exercise are shown, and hence it may be partially responsible for these considerable discrepancies.

Table 3 compares estimated and simulated trajectories in more detail showing 5th and 95th quantile for the mean of the whole set of obtained trajectories (with N=20). Because the effect of the lower number of observations on the estimated trajectories did not prove presented are only figures for 100 observations. Again, the model presented by Smets and Wouters displays perfect match between simulated and estimated trajectories. For model proposed by Lubik, there is only one small discrepancy regarding the unemployment rate ($\Delta \log U_t$ to be precise).

4 Conclusion

In this contribution I examined how the complexity of the model and the number of observations in the observables affect the results of three DSGE models, the well-known model of Smets and Wouters estimated for US data, small-scale model of Lubik focused on labour market, and rather complex model of Sheen and Wang with elements from both previously mentioned models. Analysis of the effect of complexity and number of observations was based on the simulation–estimation exercise. Starting from the given calibration, trajectories of main endogenous variables

Smets, Wouters		Lubik		Sheen, Wang	
Parameter	Value	Parameter	Value	Parameter	Value
φ	6.0144	β	0.99	b	0.6
σ_c	1.5	α	0.67	B	0.12
λ	0.6361	σ	1	$\tilde{S}'$	7.694
ξ_w	0.8087	ξ	0.7	$\tilde{\phi}_a$	0.01
σ_l	1.9423	ρ	0.1	f	0.5
ξ_p	0.6	η	0.5	η_c	1.2
ι_w	0.3243	b	0.4	η_i	1.2
ι_p	0.47	ψ	1	η_f	1.2
ψ	0.2696	κ	0.05	$\overline{\mu}^z$	1.005
Φ	0.05	$\rho_{\{A,\chi,\mu,\varepsilon\}}$	0.9	$\overline{\lambda}^d$	1.2
r_π	1.488	$\sigma_{\{A,\chi,\mu,\varepsilon\}}$	1	$\overline{\lambda}^{mc}$	1.2
ρ	0.8762			$\overline{\lambda}^{mi}$	1.2
r_y	0.0593			ξ_d	0.675
$r_{\Delta y}$	0.2347			ξ_{mc}	0.675
$\bar{\pi}$	0.7			ξ_x	0.675
$\bar{\beta}$	0.742			ρ_R	0.8
$\bar{l}$	0			r_π	1.75
$\bar{\gamma}$	0.3982			r_y	0.125
α	0.24			r_s	0
ρ_a	0.9977			r_{Δ_π}	0
ρ_b	0.5799			r_{Δ_y}	0
ρ_g	0.9957			$\rho_{\{\zeta^c,\zeta^N,\mu^z,\epsilon,\phi,\Gamma,\tilde{z}^*,\lambda^d,\lambda^{mc},\lambda^{mi},\lambda^x,\lambda^w\}}$	0.5
ρ_I	0.7165			$\sigma_{\{\zeta^c,\zeta^N,\mu^z,\epsilon,\phi,\Gamma,\tilde{z}^*,\lambda^d,\lambda^{mc},\lambda^{mi},\lambda^x,\lambda^w\}}$	1
ρ_r, ρ_p, ρ_w	0				
σ_a	0.4618				
σ_b	1.8513				
σ_g	0.6090				
σ_I	0.6017				
σ_r	0.2397				
σ_p	0.1455				
σ_w	0.2089				
μ_p, μ_w	0				
ρ_{ga}	0.51				

Table 1 Parameter setting for simulation

	Model Smets, Wouters (2007)						
	$\log H_t$	r_t	$\Delta \log P_t$	$\Delta \log Y_t$	$\Delta \log C_t$	$\Delta \log I_t$	$\Delta \log W_t$
Simulated mean	-0.83/-1.02/-1.37	1.47/1.41/1.29	0.49/0.46/0.41	0.27/0.34/0.28	0.28/0.34/0.28	0.25/0.38/0.29	0.31/0.33/0.32
Estimated mean	-0.83/-1.02/-1.37	1.47/1.41/1.29	0.49/0.46/0.41	0.27/0.34/0.28	0.28/0.34/0.28	0.25/0.38/0.29	0.31/0.32/0.32
Simulated SD	9.14/9.19/8.89	3.13/3.15/3.03	1.24/1.25/1.18	5.66/5.70/5.59	6.05/6.09/5.97	5.87/5.93/5.81	0.63/0.63/0.62
Estimated SD	9.14/9.19/8.89	3.13/3.15/3.03	1.24/1.25/1.18	5.66/5.70/5.59	6.05/6.09/5.98	5.87/5.93/5.81	0.63/0.63/0.62
	Model Lubik (2009)						
	$\Delta \log U_t$	$\Delta \log V_t$	$\Delta \log W_t$	$\Delta \log Y_t$			
Simulated mean	-0.34/-0.17/0.03	-0.84/-0.80/-0.78	0.22/0.20/0.19	-0.01/0.02/-0.12			
Estimated mean	-0.47/-0.23/-0.11	-0.78/-0.68/-0.66	0.29/0.32/0.35	0.05/0.14/0.07			
Simulated SD	2.86/2.86/2.76	2.76/2.75/2.73	1.88/1.81/1.81	2.00/1.96/1.98			
Estimated SD	2.80/2.78/2.69	2.76/2.75/2.73	1.92/1.84/1.84	1.94/1.91/1.97			
	Model Sheen, Wang (2014)						
RWHC	$\Delta \log Y_t$	U_t^a	$\Delta \log\left(\frac{W_t}{P_t}\right)$	$\Delta \log C_t$	π_t^c	R_t	$\Delta \log I_t$	$\Delta \log X_t$
Simulated mean	0.49/0.41/0.3	7.97/8.28/9.21	0.42/0.53/0.41	0.61/0.48/-0.07	-1.10/-0.17/-0.29	0.48/0.40/0.35	0.53/0.45/0.56	1.56/1.63/1.05
Estimated mean	0.47/0.39/0.35	7.97/8.28/9.21	0.52/0.45/0.55	0.49/0.41/0.39	0.42/0.53/0.41	1.75/1.69/1.59	0.57/0.48/-0.07	0.38/0.33/0.43
Simulated SD	1.54/1.49/1.51	2.89/3.01/2.92	1.15/1.24/1.19	4.26/4.61/4.72	8.04/8.05/7.18	1.43/1.47/1.47	3.12/3.13/3.18	1.71/1.66/1.45
Estimated SD	1.41/1.44/1.46	2.80/2.91/2.81	3.11/3.12/3.17	1.51/1.47/1.48	1.14/1.23/1.18	1.32/1.32/1.42	4.05/4.42/4.58	2.69/2.63/2.70
RWNOHC								
Simulated mean	0.43/0.55/0.69	5.69/5.78/5.54	0.27/0.21/0.19	0.80/1.09/1.29	1.36/-0.75/-1.59	0.51/0.60/0.75	0.59/0.75/0.93	0.94/0.58/0.79
Estimated mean	0.52/0.60/0.74	5.69/5.80/5.54	0.59/0.76/0.94	0.43/0.54/0.69	0.27/0.21/0.19	1.14/1.28/1.46	0.77/0.98/1.21	0.57/0.49/0.49
Simulated SD	2.45/2.64/2.59	4.57/4.72/4.91	2.29/2.32/2.14	3.69/3.75/3.67	10.12/10.02/10.58	2.24/2.33/2.16	4.49/4.58/4.59	1.86/1.90/1.88
Estimated SD	2.22/2.32/2.15	4.51/4.66/4.86	4.48/4.57/4.58	2.44/2.63/2.58	2.28/2.32/2.14	1.98/2.12/2.07	3.44/3.51/3.41	4.80/4.61/4.89
NWHC								
Simulated mean	0.64/0.51/0.49	6.64/10.28/11.31	0.40/1.27/1.46	1.13/0.92/0.76	3.15/2.15/2.66	0.73/0.58/0.49	0.36/0.35/0.36	1.58/2.08/2.40
Estimated mean	0.74/0.58/0.50	6.64/10.28/11.31	0.35/0.35/0.36	0.65/0.51/0.50	0.40/1.27/1.46	1.18/1.98/2.41	1.10/0.90/0.78	0.78/0.61/0.50
Simulated SD	2.89/3.00/3.05	11.71/10.90/8.72	3.36/3.35/2.53	4.71/4.85/4.33	12.46/12.95/11.64	3.55/3.71/3.50	2.49/2.64/2.62	1.76/1.62/1.61
Estimated SD	3.54/3.71/3.50	11.69/10.88/8.69	2.47/2.62/2.60	2.88/2.99/3.04	3.36/3.34/2.52	3.14/3.07/2.53	4.50/4.63/4.11	6.19/5.64/5.27
NWNOHC								
Simulated mean	0.44/0.08/1.33	14.63/7.41/8.97	2.01/1.08/0.35	0.14/-0.54/1.89	-5.67/1.07/-0.56	0.17/-0.35/1.45	0.62/0.59/1.23	0.74/1.03/0.56
Estimated mean	0.17/-0.35/1.45	14.63/7.41/8.97	0.62/0.59/1.24	0.43/0.07/1.33	2.01/1.08/0.35	2.71/1.57/1.66	0.17/-0.54/1.82	-0.07/-0.58/1.33
Simulated SD	3.93/4.22/3.67	36.77/39.05/37.26	7.80/8.71/8.12	8.90/9.90/9.39	26.64/26.99/26.54	5.65/6.16/5.59	3.10/3.19/2.86	1.47/1.47/1.42
Estimated SD	5.65/6.16/5.59	36.77/39.05/37.26	3.08/3.17/2.84	3.92/4.21/3.66	7.80/8.70/8.12	7.18/7.78/7.54	8.74/9.75/9.24	7.48/7.97/7.74

Table 2 Comparison of estimated vs. simulated data moments (100 obs. / 75 obs. / 50 obs.)

	Model Smets, Wouters (2007)						
	$\log H_t$	r_t	$\Delta \log P_t$	$\Delta \log Y_t$	$\Delta \log C_t$	$\Delta \log I_t$	$\Delta \log W_t$
Simulated mean, 5th and 95th quantile	-6.908, 8.036	-0.291, 4.034	-0.409, 1.737	-0.059, 0.803	-0.088, 0.794	-0.205, 0.845	0, 0.508
Estimated mean, 5th and 95th quantile	-6.908, 8.036	-0.291, 4.034	-0.409, 1.737	-0.059, 0.803	-0.088, 0.794	-0.205, 0.845	0, 0.508
	Model Lubik (2009)						
	$\Delta \log U_t$	$\Delta \log V_t$	$\Delta \log W_t$	$\Delta \log Y_t$			
Simulated mean, 5th and 95th quantile	-3.047, 1.896	-2.136, 0.526	-1.112, 1.591	-1.359, 1.504			
Estimated mean, 5th and 95th quantile	-3.047, 1.872	-2.136, 0.526	-1.112, 1.591	-1.359, 1.504			

Table 3 Comparison of quantiles for the mean of estimated vs. simulated trajectories with 100 observations

were simulated and used, with the various number of observations, as observables for the estimation of the model parameters. Based on my results, no significant effect of the number of observations on the estimated trajectories can be found. An indirect relation between the complexity of the model and its ability to replicate the trajectories of observables did not prove to be applicable for the examined models. The most complex model of Sheen and Wang shows more discrepancies between properties of simulated trajectories and their estimated counterparts. However, the same conclusion cannot be drawn for the case of the similarly complex model by Smets and Wouters, where almost perfect match between the properties of the investigated trajectories can be found. In addition, the presented results for Sheen and Wang model are based on only one execution of the algorithm, whilst 20 repetitions were applied for the other models. This may be responsible for at least some part of the discrepancies found in the results of Sheen and Wang model.

Acknowledgements

This work was supported by funding of specific research at Faculty of Economics and Administration, Masaryk University, project MUNI/A/0966/2017. This support is gratefully acknowledged.

References

[1] Canova, F. and Sala, L. (2009). Back to Square One: Identification Issues in DSGE Models. *Journal of Monetary Economics*, 56(4), 431-449.

[2] Guerrón–Quintana, P. A. (2010). What you match does matter: the effects of data on DSGE estimation. *Journal of Applied Econometrics*, 25(5), 774–804.

[3] Iskrev, N. (2010). Local identification in DSGE models. *Journal of Monetary Economics*, 57(2), 189–202.

[4] Koop, G., Pesaran, M. H., and Smith, R. P. (2013). On Identification of Bayesian DSGE Models, *Journal of Business & Economic Statistics*, 31(3), 300–314.

[5] Lubik, T. (2009). Estimating a Search and Matching Model of the Aggregate Labor Market. *Economic Quarterly*, 95(2), 101–120.

[6] Sheen, J. & Wang, B. Z. (2014). *An Estimated Small Open Economy Model with Labour Market Frictions*. Dynare Working Papers 35, CEPREMAP.

[7] Smets, F. & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, 97(3), 586–606.

Measuring the Efficiency of Costumer Satisfaction with DEA models

Lucie Chytilova[1]

Abstract. Nowadays, many people are buying various things, food and services at discount portals. It can be said that this is being the hit of the last few years. Slevomat is the main discount portal in the Czech Republic and it is known as very solvent. However, customer satisfaction with the obtained goods or services is varied, at least according to the portal ratings. So it was decided to take the advantage of more sophisticated tests and the knowledge of efficiency measurement to find out how customers are really satisfied. The article uses DEA models that work with ordinal data. The qualitative data are obtained on the basis of the text analysis and changed to ordinal data. Overall, the aim of the article is to measure the efficiency of customer satisfaction with Slevomat for travel services using DEA models. The obtained results give an overview of the industry and compare the results with the given server.

Keywords: DEA, text analysis, qualitative data, ordinal data, efficiency.

JEL classification: C61, D24, H43, O22, O32
AMS classification: 90C90

1 Introduction

The present time is quite materialistically oriented. People are buying various things, food and services. They are spending a lots of money. At same time, they love to save money as well. So they are trying different ways how to use discount. Nowadays, one of the most favorite ways how to save money is to use a discount portal. In the Czech Republic there are many discount portals. People may buy there almost everything - food, all kinds things (clothes, electronics etc.), vacation and so on. One of the most famous discount portal is *Slevomat* in the Czech Republic.

Slevomat[2] is a Czech company founded by T. upr, P. Barto and R. Sudov originally as a server of mass shopping. Later, Miton became a member of the company. The server was launched on April 12, 2010 as one of the first portals of its kind in the Czech Republic. Slevomat operates in 7 European countries (Czech Republic, Slovakia, Hungary, Bulgaria, Lithuania, Latvia, Estonia). Slevomat is the place where registered customers are looking for inspiration for new experiences in the form of favored offers for goods or services in several categories - travel, wine, restaurants, cosmetics, fashion and other services. Slevomat launches new offers every day. They help businesses to be seen, sell, and empty capacities. Slevomat proved to be the fastest growing company in the Czech Internet to reach an annual turnover of 1 billion crowns in less than four years since its inception. Slevomat team has about 200 people. Slevomat's principle is to negotiate with a merchant about a discount of more than 40 % on a service or product, on condition that it is at least ordered by a certain number of people. After the discount is purchased by this number of visitors to the server, the discount is activated and until now the visitors pay. After payment, visitors receive a so-called voucher containing unique code. This unique code serves the customer as a withdrawal ticket to recieve the product from merchant. Slevomat seeks to cooperate only with competent merchants. Competence is verified by customer reviews (star and word ratings). The question is, if evaluation by stars (5 stars mean a satisfaction and 0 star means discontent) that most people are looking for, is the best rating.

Measurement of customer satisfaction in Slevomat is based on stars, which may not be as accurate as mentioned above. In this respect, customer satisfaction can be understood as the efficiency of the service, which can be seen for example in the article [1]. And since Slevomat, along with star rating, also includes the verbal evaluation (something as a word rating), it allows us to explore more deeply the customer satisfaction. Basic Data Envelopment Analysis model (DEA) is a models that solves technical efficiency and works with the quantitative data. There are other types of DEA models that work with qualitative data, can be seen in the work [6]. So, on the basis of the analysis of the text, it is possible to analyze the efficiency / satisfaction of customers with one travel agency (the rule of homogeneity must be met). More specifically, the goal of the article is to measure the efficiency of customer satisfaction with Slevomat for travel services by use of DEA models. The obtained results provided an overview of the field. Plus, there had been compared with the results given by the server.

[1]VSB-TU Ostrava, Economic faculty, Department of System Engineering, Sokolska tr. 33, Ostrava, Czech Republic, lucie.chytilova@vsb.cz

[2]www.slevomat.cz

The remainder of this paper is organized as follows. Next section discuss the methodology of both DEA models. Section three introduces the used data set. Section four presents analysis of results and finally, section five provides some final considerations and conclusions.

2 Data envelopment analysis (DEA) and imprecise DEA (IDEA)

Data envelopment analysis (DEA) is non-parametric approach. It measurs the relative efficiency of decision making units (DMUs) when the data for multiple inputs and outputs are known exactly. Suppose, there is a set of T peer DMUs, (DMU$_k$ for $k = 1, ..., T$), with multiple inputs x_{ik} ($i = 1, ..., R$) and multiple outputs y_{jk} ($j = 1, ..., S$), respectively. Also, u_i ($i = 1, ..., R$) and v_j ($j = 1, ..., S$) be the weight of i^{th} input and j^{th} output, respectively. Mathematically, the relative efficiency score of DMU$_k$ 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. [2] proposed the very first DEA model - CCR ratio model for measuring the efficiency score of under evaluation unit, DMU$_Q$ where $Q \in \{1, ..., T\}$:

$$\begin{array}{lll} & \max e_Q = \frac{\sum_{j=1}^{S} v_j y_{jQ}}{\sum_{i=1}^{R} u_i x_{iQ}}, & \\ \text{s.t.} & \frac{\sum_{j=1}^{S} v_j y_{jk}}{\sum_{i=1}^{R} u_i x_{ik}} \leq 1, & k = 1, ..., T, \\ & u_i > 0, & i = 1, ..., R, \\ & v_j > 0, & j = 1, ..., S. \end{array} \tag{2}$$

This model (2) transformed into the following CCR multiplier problem:

$$\begin{array}{lll} & \max e_Q = \sum_{j=1}^{S} v_j y_{jQ}, & \\ \text{s.t.} & \sum_{i=1}^{R} u_i x_{iQ} = 1, & \\ & \sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} \leq 0, & k = 1, ..., T, \\ & u_i > 0, & i = 1, ..., R, \\ & v_j > 0, & j = 1, ..., S. \end{array} \tag{3}$$

When x_{ik} (for any i) and y_{jk} (for any j) are imprecise and unknown decision variables as bounded and ordinal data, see Cook et al. [3], [4], model (3) becomes a non-linear and non-convex program, and it is called imprecise DEA (IDEA), see e.g. Zhu [7]. The bounded data expressed as:

$$\underline{y_{jk}} \leq y_{jk} \leq \overline{y_{jk}} \text{ and } \underline{x_{ik}} \leq x_{ik} \leq \overline{x_{ik}}, \tag{4}$$

for $j \in BO$ and $i \in BI$, where y_{jk} and x_{ik} are the lower bounds, and y_{jk} and x_{ik} are the upper bounds. BO and BI are the associated sets. These sets contain bounded outputs and inputs, respectively.

The weak ordinal data expressed as:

$$y_{jk} \leq y_{jl} \text{ and } x_{ik} \leq x_{il} \ \ \forall k \neq l \tag{5}$$

for $j \in$ DO and $i \in$ DI, or to simplify the presentation,

$$y_{j1} \leq y_{j2} \leq ... \leq y_{jl} \leq ... \leq y_{jn} \ \ (j \in DO), \tag{6}$$

$$x_{i1} \leq x_{i2} \leq ... \leq x_{il} \leq ... \leq x_{in} \ \ (i \in DI),$$

where DO and DI represent the associated sets containing weak ordinal outputs and inputs, respectively.

The strong ordinal data expressed as:

$$y_{j1} < y_{j2} < ... < y_{jl} < ... < y_{jn} \ \ (j \in SO), \tag{7}$$

$$x_{i1} < x_{i2} < ... < x_{il} < ... < x_{in} \ \ (i \in SI),$$

where SO and SI represent the associated sets containing strong ordinal outputs and inputs, respectively.

Model (3) involving (4) - (7) then changes into the following model:

$$\begin{array}{ll} & \max e_Q = \sum_{j=1}^{S} v_j y_{jQ}, \\ \text{s.t.} & \sum_{i=1}^{R} u_i x_{iQ} = 1, \\ & \sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} \leq 0, \quad k = 1, 2, ..., T, \\ & (x_{ik}) \in \theta_i^-, \qquad\qquad\qquad\quad i = 1, 2, ..., R, \\ & (y_{jk}) \in \theta_j^+, \qquad\qquad\qquad\quad j = 1, 2, ..., S, \\ & u_i \geq 0, v_j \geq 0, \end{array} \tag{8}$$

where $(x_{ik}) \in \theta_i^-$ and $(yjk) \in \theta_j^+$ represents any of or all of (4) - (7) sets. The model (8) with the (4) - (6) sets is solved by the standard linear DEA models via concerning the bounded and ordinal data into exact data, Zhu [7].

The model (8) when θ_i^- and θ_j^+ are in forms of (6) set and obtain a set of optimal solutions y_{jk}^* and x_{ik}^* with the optimal e_Q^* is the following:

$$\begin{array}{ll} & e_Q^* = \max \sum_{j \in DO} v_j y_{jQ} + \sum_{j \notin DO} v_j y_{jQ}, \\ \text{s.t.} & \sum_{i \in DI} u_i x_{iQ} + \sum_{i \notin DI} u_i x_{iQ} = 1, \\ & \sum_{j \in DO} v_j y_{jk} + \sum_{j \notin DO} v_j y_{jk} - \sum_{i \in DI} u_i x_{ik} \\ & - \sum_{i \notin DI} u_i x_{ik} \leq 0, \\ & k = 1, 2, ..., T, \\ & u_i \geq 0, v_j \geq 0, \quad i = 1, 2, ..., R, j = 1, 2, ..., S, \\ \text{where} & 0 \leq y_{j1}^* \leq y_{j2}^* \leq ... \leq y_{jl}^* \leq ... \leq y_{jn}^* \leq M \ \ (j \in \textit{DO}), \\ & 0 \leq x_{i1}^* \leq x_{i2}^* \leq ... \leq x_{il}^* \leq ... \leq x_{in}^* \leq M \ \ (i \in \textit{DI}). \end{array} \tag{9}$$

Where y_{jk}^* and x_{ik}^* are obtained a set of optimal solutions with the optimal e_Q^* and M is sufficiently large, more accurate information are at articles by Zhu [7]. Note that in this case, model is no longer a non-linear program, but a linear CCR model.

3 Data set

There are 150 observations/DMUs in the analysis. More specifically, there are 150 ratings (verbal evaluations) for a one-day trip abroad with travel agency *Cesty Evropou*. The data were collected directly from the server *Slevomat*. The ratings were taken for the time period from 2016 to 2018. Table 1 provides a statistical description of the data.

	stars	positives (+)	negatives (-)
sum	695	195	47
max	5	4	3
min	1	1	1
std.dev	0.77	0.83	0.64
average	4.63	1.86	1.34

Table 1 Statistical description of the data

Table 1 shows the sum, maximum, minimum, standard deviation and average value. These statistics are listed for the stars, positives and negatives texts - verbal evaluations. It can be seen that starts have the maximum in five stars (satisfied clients) and the minimum is one star (dissatisfied client; there is no one-day trip with rating of zero stars). According to the sum and average (150 observations), clients were primarily satisfied with the one-day trip (4.63 stars). The positives and negatives texts are based on a textual analysis. From Table 1 it can be seen that positive comments had prevail in this analysis. So again, it can be seen that clients were satisfied.

Positive and negative text from the text analysis was modified for need of DEA models. The input-oriented DEA model was used. It means that the outputs should be maximized, i.e. the positive text was defined as the output. The input side was defined as negative text, i.e. the variables should be minimized.

More specifically, the positive text has been divided into four categories - organization, guide, transport and another positives. The first three categories were the most represented comments and the last one was left for another positive comments. The transformation of the text had been done as follows. The output always had a value of one in each category, and if another positive aspect has been found in the category, another value of one

had been added. For example, in the beginning the vector of outputs for the customer one (DMU01) was $(1, 1, 1, 1)$. After the textual analysis, where it was found out that the customer said that the guide was very erudite and the drivers were professional, the output vector had changed. So, the final output vector for this DMU is defined as $(1, 2, 2, 1)$. Inputs were define similarly. Inputs were divided into three categories - communication with travel agency, organization of the trip and another negatives. Again, in general, the inputs have a value of 1, and if there was a negative mention in the text, the number one was added. An example is rating of customer fifteen (DMU15), where the customer complained of poor communication with the travel agency. The final vector of inputs for this DMU is $(2, 1, 1)$.

For better clarity of previous paragraph, Table 2 introduces a part of the data file, more precisely the first 20 units of measurement. DMU01 and DMU15 from previous paragraph are seen in Table 2. Table 2 shows the verbal description of the plus and minus that had been subdivided into the subgroups (see above). Then the number of stars is given, i.e. the total rating, the number of the plus and minus, their difference and finally the efficiency score is there as well.

DMU	positives in words	negatives in words	stars	+	-	difference	efficiency score
DMU01	driver, guide		5	2	0	2	0.600
DMU02			5	0	0	0	0.100
DMU03	guide, organization		5	2	0	2	0.200
DMU04	location		5	1	0	1	0.800
DMU05	location		5	1	0	1	0.800
DMU06	bus, driver		5	2	0	2	0.800
DMU07		guide	4	0	1	-1	0.600
DMU08			5	0	0	0	0.600
DMU09	guide, location		5	2	0	2	0.600
DMU10	guide		5	1	0	1	0.600
DMU11			5	0	0	0	0.400
DMU12	location, transportation, guide		5	3	0	3	0.400
DMU13	time schedule, driver, guide		5	3	0	3	0.400
DMU14	location, driver	bus, guide	4	2	1	1	0.400
DMU15		organization	3	0	1	-1	0.400
DMU16	location	bus	5	1	1	0	0.400
DMU17	location, organization, bus	guide	4	3	1	2	0.300
DMU18	organization		5	1	0	1	0.200
DMU19	organization		5	1	0	1	0.200
DMU20	guide	organization	4	1	1	0	0.200
...							

Table 2 Part of the data file

4 Results and discussions

Three analyzes of all the results had been done and compared. Averages of all results are shown in Table 3.

	stars	positives	negatives	differences	efficiency score
sum	695	195	47	148	61.35
max	5	4	3	4	1
min	1	0	0	-3	0.05
std.dev	0.77	1.1	0.65	1.37	0.24
average	4.63	1.3	0.31	0.99	0.41

Table 3 Statistical description of all analysis

The first analysis concerned only the ratings by stars. Based on the average value of the stars, which is 4.63, it can be assumed (if the 5 stars mean most satisfied and the 0 stars mean least satisfied customer) that the customers were generally satisfied with the one-day trip, without any big problem.

The second analysis was based on the difference between positives and negatives. The principle was the use of the textual analysis. A distinction was made between positive and negative references in the text. If the distinction was positive, one can assume that the customer was more than satisfied. The more positive the difference between positives and negatives, the customer is more satisfied. In the case of a negative difference, we are talking about dissatisfaction. If the value of the difference is zero, it can be assumed that the customer was satisfied, but he/she did not received anything extra then it was expected. So he/she had no reason to comment on it or that the negatives were balanced by the positive. Table 2 shows that the positive was greater than the negative. The difference between positives and negatives is 148. Generally speaking, customers are satisfied. More precisely, the average value, which is 0.99, indicates that the satisfaction with the one-day trip always prevails.

The third analysis was based on the IDEA models. As already mentioned, DEA models calculates technical efficiency, but based on input and output data, it can be said that technical analysis can equal customer satisfaction. If the customer is 100 % satisfied, the value is one. If a customer has already found some problem or he/she was not satisfied fully, the value is between 0 and 0.99. According to the IDEA method, there had been just three customers totally satisfied. Another 50 customers had the satisfaction score (the efficiency score) in interval 0.99 - 0.6. These customers probably were satisfied about the one day trip, but they did not get anything extraordinarily superior (in their eyes). 80 customers had the satisfaction score in interval 0.59 - 0.2. These customers are close to average. It means that they get what they wanted probably and the one day trip was good. And the rest of the customers, with the satisfaction score in interval 0.19 - 0, was not satisfied at all.

The last analysis should be the most comprehensive. The reason is the largest database of inputs and outputs. Table 3 shows that this analysis is most pessimistic. The average value of an effective score is 0.41. This value means that customers are satisfied in half. It has to be said that this can be influenced by the fact that some customers did not comment on the given categories, so it can be expected that the score is not 1. However, the value of 0.41 is quite small and from a classical psychological point of view it can be said that customers were not extremely attentive to the service, but they have met their expectations.

5 Conclusion

The aim of the article is to measure the efficiency of customer satisfaction with Slevomat for travel services using the score of Slevomat, positive and negative text and IDEA models.

The results of the first two analyzes are similar. These analyzes generally refer to satisfied customers. The latest analysis gives much more varied results, but can not be interpreted exactly as the classic DEA method. The value one means that the rating is effective, but due to the data it also means that the customer satisfaction was extreme. In other words, the customer received something above average, something above expectations. Values around the average can then be classified as classically satisfied by customers.

Therefore, the goal of the article has been fulfilled, but it is still necessary to look at the results of the results and find better data bases.

Acknowledgements

Supported by the grant No. SP2018/146 of the Student Grants Support at VSB-TU Ostrava.

References

[1] Bayraktar, E. et al. (2012). Measuring the efficiency of customer satisfaction and loyalty for mobile phone brands with DEA. *Expert Systems with Applications* (39), 99–106.

[2] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research* (6), 429–444.

[3] Cook, W.D., Kress, M. & Seiford, L.M. (1993). On the use of the ordinal data in Data Envelopment Analysis. *Journal of the Operational Research Society* (44), 133–140.

[4] Cook, W.D., Kress, M. & Seiford, L.M. (1996). Data envelopment analysis in the presence of both quantitative and qualitative factors. *Journal of the Operational Research Society* (47), 945–953.

[5] Cooper, W. W., Seiford, L. M. & Zhu, J. (2004). *Handbook on Data Envelopment Analysis* (1st edition). Boston: Springer.

[6] Inuiguchi, M., Mizoshita, F. (2012). Qualitative and quantitative data envelopment analysis with interval data. *Annals of Operations Research* (1), 189–220.

[7] Zhu, J. (2002). Imprecise data envelopment analysis (IDEA): A review and improvement with an application. *European Journal of Operational Research* (144), 513–529.

Alternative approaches for efficiency evaluation in multi-stage serial DEA models

Josef Jablonský[1]

Abstract. Traditional data envelopment analysis (DEA) models have been proposed as a tool for relative efficiency and performance evaluation of the set of homogeneous units that transforms multiple inputs into multiple outputs in one stage. Reality of often more complex and the process under evaluation may consists of several consecutive serial stages. Multi-stage models assume that all or at least a certain part of outputs of a stage *s* serve as the inputs of the next stage (*s*+1) and the outputs of the last stage of the model are considered as the final outputs of the production process. It is possible evaluating efficiency of each stage of the model independently but there are proposed several models trying to evaluate the overall efficiency of the multi-stage production process. Kao and Hwang [9] and Chen et al. [3] models belong to the most known ones. The paper compares traditional multi-stage DEA models with alternative approaches based on multi-objectiveoptimization methodology. A minimax, weighted sum of deviations and lexicographic principles are applied. The results are illustrated on a two-stage model with a real data set with 20 units.

Keywords: Data envelopment analysis, multi-stage model, efficiency, ranking.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Conventional DEA models, firstly introduced by Charnes et al. [2] and further extended by Banker et al. [1] and by many others deal with efficiency analysis of single-stage processes. However, reality is usually more complex and the production process under consideration can be defined as a multi-stage process. It can be analyzed using network DEA models that have been introduced by Fare and Grosskopf [5] at first. A simplest case of multi-stage processes is the one that consist of two serial sub-processes. Kao and Hwang [9] and Chen et al. [3] are two first DEA models for efficiency evaluation of two-stage serial models but they can be easily generalized for a case of multi-stage systems. They are based on traditional definition of efficiency as the weighted sum of outputs divided by the weighted sum of inputs. A slack-based model (SBM) for two-stage processes is proposed by Jablonský [8]. Two-stage systems and their evaluation of efficiency using DEA models is still subject of research among many researchers. Kao [10] is an interested review of two-stage DEA models. A similar study about two-stage DEA models is the main topic of the paper Halkos et al. [6]. An importance and attractiveness of the research in this field is illustrated by the latest studies Despotis et al. [4] or Li et al. [11]. They discuss the latest developments in network DEA models and propose original models extending the traditional ones.

The aim of this paper is to discuss the possibility of using alternative approaches for efficiency and performance analysis in multi-stage serial production systems. The models proposed further in our paper are based on the methodology of multi-objective optimization and work in two main stages. In the first stage the efficiency scores of individual stages are derived using traditional DEA models. The next stage consists in an application of multi-objective optimization principles – lexicographic optimization, minimization of maximum deviation from ideal values (maximum efficiency scores from the first stage), and minimization of the sum of deviations. This process lead to optimal weights of the inputs, final outputs and all intermediate variables that minimize the given metrics and derives efficiency scores of particular stages and/or overall efficiency score that allows complete ranking of the units under evaluation.

The paper is organized as follows. Section 2 contains basic definitions and formulation of traditional multi-stage models. Section 3 introduces original multi-stage DEA models based on multi-objective optimization methodology. The results of the model are illustrated on the two-stage model with a real data set – efficiency evaluation of 20 Czech banks. Final section discusses given results and possibilities of a future research in this field.

[1] University of Economics, Prague, Faculty of Informatics and statistics, Department of Econometrics, W. Churchill Sq. 4, Praha 3, Czech Republic, e-mail: jablon@vse.cz

2 Multi-stage serial DEA models

Let us assume a multi-stage serial production process with S stages. It is graphically illustrated on Figure 1. The first stage of this process spends m inputs and produces p_1 outputs. For simplicity, we suppose that all outputs of the first stage serve as the inputs of the second stage (in general, some of them may leave production process). The second stage produces p_2 outputs that enter the third stage as its inputs. In general, the stage $s = 2,\ldots,S\text{-}1$ has p_{s-1} inputs and p_s outputs. Let us denote for the DMU$_i$: x_{ij}, $i = 1,\ldots,n$; $j = 1,\ldots,m$, the value of the j-th input of the first stage, y_{ik}, $i = 1,\ldots,n$; $k = 1,\ldots,r$, the value of the k-th final output of the last stage, and $z^s_{il_s}$, $i = 1,\ldots,n$; $l_s = 1,\ldots,p_s$, $s = 1,...,S-1$ the value of the l_s-th output of the s-th stage and at the same time the value of the l_s-th input of the (s+1)-th stage. Individual efficiency scores of the first and the last stages θ^1_q and θ^S_q are defined as follows:

$$\theta^1_q = \frac{\sum_{l_1=1}^{p_1} w^1_{l_1} z^1_{ql_1}}{\sum_{j=1}^{m} v_j x_{qj}}, \qquad \theta^S_q = \frac{\sum_{k=1}^{r} u_k y_{qk}}{\sum_{l_{S-1}=1}^{p_{S-1}} w^{S-1}_{l_{S-1}} z^{S-1}_{ql_{S-1}}}, \tag{2}$$

where $w^s_{il_s}$ $l_s = 1,\ldots,p_s$, $s = 1\ldots.,S-1$, are the weights of the intermediate variables in the s-th stage. The overall efficiency score θ^0_q for the DMU$_q$ is the weighted sum of final outputs divided by the weighted sum of the inputs of the first stage, i.e.

$$\theta^0_q = \frac{\sum_{k=1}^{r} u_k y_{qk}}{\sum_{j=1}^{m} v_j x_{qj}}. \tag{3}$$

It is clear that the overall efficiency score is the product of efficiency scores of all particular stages, i.e. $\theta^0_q = \theta^1_q ... \theta^S_q$ holds.

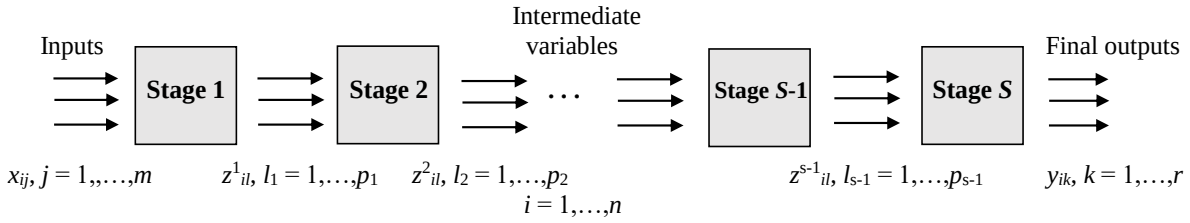


Figure 1: *Multi-stage serial model*

Kao and Hwang model [9] was originally proposed for two-stage serial models but can easily be generalized for a multi-stage case. This generalized model maximizes overall efficiency score θ^0_q of the unit under evaluation under the assumption that efficiency scores in all other stages for all other units of the set do not exceed 1. This model is not linear in its objective function but can be easily transformed into a linear program. The linearized version of this model is below (the model assumes $S \geq 3$; for two-stage models the second set of constraints must be removed):

$$\begin{aligned}
&\text{Maximize} && \theta^{*KH}_q = \sum_{k=1}^{r} u_k y_{qk} && \\
&\text{Subject to} && \sum_{l_1=1}^{p_1} w^1_{l_1} z^1_{il_1} \leq \sum_{j=1}^{m} v_j x_{ij}, && i = 1,\ldots,n, \\
& && \sum_{l_s=1}^{p_s} w^s_{l_s} z^s_{il_s} \leq \sum_{l_{s-1}=1}^{p_{s-1}} w^{s-1}_{l_{s-1}} z^{s-1}_{il_{s-1}}, && i = 1,\ldots,n,\ s = 2,\ldots,S-1, \\
& && \sum_{k=1}^{r} u_k y_{ik} \leq \sum_{l_{S-1}=1}^{p_{S-1}} w^{S-1}_{l_{S-1}} z^{S-1}_{il_{S-1}}, && i = 1,\ldots,n,
\end{aligned} \tag{4}$$

$$\sum_{j=1}^{m} v_j x_{qj} = 1,$$
$$v_j > 0, w_l > 0, u_k > 0.$$

The ideal (maximum) efficiency score of the unit DMU_q in all stages can be derived using conventional input-oriented DEA models with constant returns to scale technology. The model for a general stage $s = 2,\ldots,S\text{-}1$ is as follows (the model for deriving ideal efficiency scores for the first and the last stages are slightly modified):

$$\begin{aligned}
&\text{Maximize} && \theta_q^{*s} = \sum_{l=1}^{p_s} w_l^s z_{ql}^s \\
&\text{subject to} && \sum_{l=1}^{p_s} w_l^s z_{il}^s \le \sum_{l=1}^{p_{s-1}} w_l^{s-1} z_{il}^{s-1}, \quad i = 1,\ldots,n, \\
& && \sum_{l=1}^{p_{s-1}} w_l^{s-1} z_{ql}^{s-1} = 1, \\
& && w_l^s > 0, w_l^{s-1} > 0.
\end{aligned} \tag{5}$$

3 Alternative models for efficiency evaluation in multi-stage processes

The main aim of the paper is to verify the applicability of alternative approaches for measuring efficiency in case of multi-stage models. Three models based on basic multi-objective optimization principles will be taken into account – lexicographic optimization, minimization of the sum of deviations from ideal values, and minimax principle.

3.1 Lexicographic optimization

Let us assume that the decision maker is able to discriminate among levels of efficiencies in all particular stages. Without the loss of generality we can assume that the efficiency score of the first stage is the most important, the second stage has the second importance, etc. Under this assumption we can solve the following lexicographic optimization problem:

$$\begin{aligned}
&\text{lex max} && (\theta_q^1, \theta_q^2, \ldots, \theta_q^S) \\
&\text{subject to} && \sum_{l_1=1}^{p_1} w_{l_1}^1 z_{il_1}^1 \le \sum_{j=1}^{m} v_j x_{ij}, && i = 1,\ldots,n, \\
& && \sum_{l_s=1}^{p_s} w_{l_s}^s z_{il_s}^s \le \sum_{l_{s-1}=1}^{p_{s-1}} w_{l_{s-1}}^{s-1} z_{il_{s-1}}^{s-1}, && i = 1,\ldots,n,\ s = 2,\ldots,S-1, \\
& && \sum_{k=1}^{r} u_k y_{ik} \le \sum_{l_{S-1}=1}^{p_{S-1}} w_{l_{S-1}}^{S-1} z_{il_{S-1}}^{S-1}, && i = 1,\ldots,n, \\
& && v_j > 0, w_l > 0, u_k > 0.
\end{aligned} \tag{6}$$

Let us denote (**v***, **w**^{1}*, **…**, **w**^{S}*, **u***) vectors of optimal weights computed by model (6). The efficiency scores of the first and the last stages and overall efficiency score using these optimal weights are as follows (efficiency scores of intermediate stages are computed in the same way):

$$\theta_q^1 = \frac{\sum_{l_1=1}^{p_1} w_{l_1}^* z_{ql_1}^1}{\sum_{j=1}^{m} v_j^* x_{qj}}, \quad \theta_q^S = \frac{\sum_{k=1}^{r} u_k^* y_{qk}}{\sum_{l_{S-1}=1}^{p_{S-1}} w_{l_{S-1}}^* z_{ql_{S-1}}^{S-1}}, \qquad \theta_q^0 = \theta_q^1 \ldots \theta_q^S. \tag{7}$$

Results (7) allow ranking of the DMUs easily according to their overall efficiency scores θ_i^0, $i = 1,\ldots,n$.

3.2 Weighted sum of deviations

The model introduced below minimizes the weighted sum of deviations of efficiency scores of all stages from their ideal values obtained by models (5).

$$
\begin{aligned}
\text{minimize} \quad & \sum_{s=1}^{S} \alpha_s d_s \\
\text{subject to} \quad & \sum_{l_1=1}^{p_1} w_{l_1}^1 z_{il_1}^1 \le \sum_{j=1}^{m} v_j x_{ij}\,, && i = 1,\ldots,n, \qquad (8)\\
& \sum_{l_s=1}^{p_s} w_{l_s}^s z_{il_s}^s \le \sum_{l_{s-1}=1}^{p_{s-1}} w_{l_{s-1}}^{s-1} z_{il_{s-1}}^{s-1}, && i = 1,\ldots,n,\ s = 2,\ldots,S-1,\\
& \sum_{k=1}^{r} u_k y_{ik} \le \sum_{l_{S-1}=1}^{p_{S-1}} w_{l_{S-1}}^{S-1} z_{il_{S-1}}^{S-1}, && i = 1,\ldots,n,\\
& \sum_{l_1=1}^{p_1} w_{l_1}^1 z_{ql_1}^1 + d_1 = \theta_q^{1*} \sum_{j=1}^{m} v_j x_{qj}\,,\\
& \sum_{l_s=1}^{p_s} w_{l_s}^s z_{ql_s}^s + d_s = \theta_q^{s*} \sum_{l_{s-1}=1}^{p_{s-1}} w_{l_{s-1}}^{s-1} z_{ql_{s-1}}^{s-1}, && s = 2,\ldots,S-1,\\
& \sum_{k=1}^{r} u_k y_{qk} + d_S = \theta_q^{S*} \sum_{l_{S-1}=1}^{p_{S-1}} w_{l_{S-1}}^{S-1} z_{ql_{S-1}}^{S-1},\\
& v_j > 0, w_l > 0, u_k > 0.
\end{aligned}
$$

In model (8), d_s ($s = 1,\ldots,S$) are deviational variables that express the distance of the efficiency scores θ_q^s from their theoretically maximum values. α_s ($s = 1,\ldots,S$), $\sum \alpha_s = 1$ are the weights expressing the importance of deviational variables in the objective function. It is clear that $\alpha_i = 1$ leads to the model that maximizes efficiency score of the i-th stage but its result need not be identical with results of models (5). The weights derived by this model may be used for calculation of efficiency scores of all stages and overall efficiency scores using (7).

3.3 Minimax model

The minimax model is just a modification of model (8). Instead of minimization of the sum of deviations a maximum deviation is minimized. In model (8) its objective function is replaced by the maximum deviation D and the set of constraints is extended by:

$$
\begin{aligned}
& d_s \le D, && s = 1,\ldots,S, \qquad (9)\\
& d_s \ge 0.
\end{aligned}
$$

As in the previous model the weights computed by this model are used for calculation of particular and overall efficiency scores which allow ranking of the units under consideration.

4 Numerical illustration

The results of the above formulated models are illustrated on efficiency evaluation of 20 banks operating on the Czech financial market. It is rather an illustrative example with a real background. The data set comes from Jablonský [7] and originally is taken from public financial statements of the banks It is presented in Table 1. The first stage evaluates production efficiency and the second stage profit efficiency. The following inputs, intermediate variables and final outputs are taken into consideration.

Inputs:
- Equity of the bank in millions of CZK,
- Number of full time employees (FTE).

Intermediate variables (outputs of the first stage and inputs of the second one):
- Deposits in millions of CZK,
- Credits in millions of CZK.

Only final output:
- Profit in millions of CZK.

The problem is expressed as a two-stage model, i.e. in our notation above $S = 2$. The models (4), (6), and (8) presented in previous sections are formulated in a general way for $S \geq 3$. For the case of two-stage models they can be easily modified by omitting of the second set of constraints in each model.

• Bank	Equity	FTE	Credits	Deposits	Profit
CS	14 014	10 163	416 854	530 101	14 317
Citibank	673	900	125 062	169 425	14 310
ČSOB	3 635	6 420	209 172	568 199	13 572
GE Money	775	2 290	97 262	109 942	3 851
Hypotecní Bank	157	476	145 070	455	2 288
ING CZ	84	123	19 169	89 211	821
KB	6 556	7 883	334 834	441 285	14 417
LBBW Bank	727	350	19 161	18 929	50
Raiffeisenbank	989	528	152 663	125 936	2 320
UniCredit	1 578	265	172 070	174 373	3 473
Volksbank CZ	186	726	39 147	30 155	345
Wustenrot	380	200	31 978	35 956	341
Commerzbank	42	371	40 162	19 950	343
Czech Export B	39	144	59 856	58 690	217
Czech-Moravian	171	218	20 776	27 076	1 053
PPF Bank	44	154	18 655	36 332	804
Blue pyramid	399	356	49 030	69 119	1 026
Raiffeisen SS	56	257	39 578	76 160	830
CS SS	420	250	44 307	97 540	1 464
Wustenrot hyp	17	33	11 261	2 259	90

Table 1 Data set – efficiency evaluation of Czech banks

The results obtained by presented models for all DMUs are included in Table 2. All results are computed under the assumption of constant returns to scale returns to scale technology. Assumptions of other technologies can be ensured by small changes of the models. Table 2 contains the following information:

- Column (1) – name of the bank.
- Columns (2) to (4) – efficiency scores of the first and second stages, and overall efficiency scores – three banks are efficient in the first stage (ING CZ, UniCredit and Czech Export) and two in the second stage (Citibank and Hypotecni Bank). Citibank is only DMU overall efficient.
- Column (5) – efficiency scores obtained by Kao and Hwang model (4).
- Column (6) – overall efficiency scores derived using of the optimal weights of the model (8) with $\alpha_1 = \alpha_2 = 0.5$.
- Column (7) – overall efficiency scores computed using optimal weights of the minimax model (9).

The results show that there are no significant differences among all presented models. There are few rank reversals of the banks but all model identify as the most efficient Citibank.

5 Conclusions

The presented three models for efficiency and performance evaluation of serial multi-stage processes can be considered as a possible alternative to traditional network DEA models. The numerical example shows that their results are more or less similar to traditional models. There is almost a perfect correlation between the results of Kao and Hwang model on one hand and the other models on the other hand (correlation coefficient greater than 0.99 in all pairs of results).

The advantage of the proposed models consists in their parametrization which allows a more detailed analyzes of level of efficiency of the units under evaluation. The relation of the presented models to conventional network and/or single-stage models was verified on an example only. A general relation should be further investigated which could be an excellent issue for a future research.

Bank	1st stage	2nd stage	Overall	Kao-Hwang	WSD	Minimax
(1)	(2)	(3)	(4)	(5)	(6)	(7)
CS	0.1010	0.3170	0.0886	0.0314	0.0314	0.0312
Citibank	0.3764	1.0000	1.0000	0.3764	0.3764	0.3764
ČSOB	0.1401	0.5671	0.1756	0.0509	0.0510	0.0486
GE Money	0.1088	0.4038	0.2337	0.0427	0.0415	0.0420
Hypotecní Bank	0.7290	1.0000	0.6854	0.1234	0.1208	0.1166
ING CZ	1.0000	0.3743	0.4597	0.1590	0.1590	0.1529
KB	0.1124	0.3853	0.1150	0.0429	0.0429	0.0428
LBBW Bank	0.1117	0.0298	0.0090	0.0030	0.0030	0.0030
Raiffeisenbank	0.6004	0.2006	0.2763	0.0936	0.0867	0.0876
UniCredit	1.0000	0.2255	0.8243	0.2255	0.2255	0.2255
Volksbank CZ	0.1371	0.1228	0.0872	0.0139	0.0115	0.0112
Wustenrot	0.3522	0.1092	0.1072	0.0363	0.0363	0.0361
Commerzbank	0.6231	0.1649	0.3841	0.0611	0.0485	0.0485
Czech Export	1.0000	0.0416	0.2617	0.0416	0.0416	0.0416
Czech-Moravian	0.2518	0.4580	0.3038	0.1139	0.1139	0.1133
PPF Bank	0.5630	0.3767	0.8594	0.1415	0.1347	0.1347
Blue pyramid	0.3680	0.1829	0.1813	0.0656	0.0658	0.0656
Raiffeisen SS	0.9037	0.1833	0.6971	0.1109	0.1036	0.0880
CS SS	0.5925	0.2888	0.3683	0.1270	0.1270	0.1238
Wustenrot hyp	0.8016	0.2648	0.2490	0.0661	0.0610	0.0610

Table 2 Efficiency scores obtained by DEA models

Acknowledgements

The research is supported by the Czech Science Foundation, project no. 16-01821S - *Models for efficiency evaluation, benchmarking and resource allocation in public sector*.

References

[1] Banker, R.D., Charnes, A. & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.
[2] Charnes, A., Cooper, W.W. & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
[3] Chen, Y., Liang, L. & Zhu, J. (2009). Equivalence in two-stage DEA approaches. *European Journal of Operational Research*, 193(2), 600-604.
[4] Despotis, D., Koronakos, G. and Sotiros, D. (2016). Composition versus decomposition in two-stage network DEA: A reverse approach. *Journal of Productivity Analysis*, 45(1), 71-87.
[5] Fare, R. & Grosskopf, S. (2000). Network DEA. *Socio-Economic Planning Sciences*, 34(1), 35-49.
[6] Halkos, G.E., Tzeremes, N.G. and Kourtzidis, S.A. (2014). A unified classification of two-stage DEA models. *Surveys in Operations Research and Management Science*, 19(1), 1-16.
[7] Jablonský, J. (2012). Data envelopment analysis models with network structure. In: *Mathematical Methods in Economics 2012*. Karviná, pp. 367–372.
[8] Jablonský, J. (2018). Ranking of countries in sporting events using two-stage data envelopment analysis models: a case of Summer Olympic Games 2016. *Central European Journal of Operations Research*, https://doi.org/10.1007/s10100-018-0537-8.
[9] Kao, C. & Hwang, S.N. (2008). Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *European Journal of Operational Research*, 185(1), 418-429.
[10] Kao, C. (2014). Network data envelopment analysis: A review. *European Journal of Operational Research*, 239(1): 1-16.
[11] Li, H., Chen, C., Cook, W.D., Zhang, J. & Zhu, J. (2018). Two-stage network DEA: Who is the leader? *Omega*, 74(1), 15-19.

Stakelberg Game in Emergency Service System Reengineering

Jaroslav Janáček[1], Marek Kvet[2]

Abstract. An emergency service system design is usually worked up by a system administrator, who acts on behalf of the public and tries to minimize disutility perceived by an average user. As distribution of demands for service develops in time and space, the originally determined center locations will cease to suit both serviced population and providers and reengineering of the system is to be performed. In some systems, the administrator is responsible for the reengineering. In other systems, the administrator only defines some rules, under which a service provider is allowed to relocate his service centers. In the second mentioned emergency health care system the profit of a provider is proportional to transportation performance necessary for the demand satisfaction. This paper deals with the latter case, when partial reengineering is suggested by the private service providers running considerable portions of the current service centers. The providers try to maximize their profit subject to the system administrator's rules. We model the providers' behavior and the process of reengineering as an instance of Stackelberg's game.

Keywords: Location, Emergency Medical Service, System Reengineering, Stackelberg's Game of Providers.

JEL Classification: C61
AMS Classification: 90C27

1 Introduction

This contribution was motivated by a special case of emergency service system reengineering, when the decisions on change of service center deployment is not made by system administrator, but a considered service provider, who acts subject rules given by the administrator. Originally, the emergency system is designed with the objective to minimize the average or total disutility perceived by the users [1], [3], [8]. The optimal deployment of service centers for such type of system can be obtained by exact or approximate solving of the weighted *p*-median problem modelled either by the location-allocation or radial formulations [4], [5], [6], [12]. The administrator also supervises dispatching of emergency vehicles to individual users' demands in the way that each user demand is served from the nearest available service center. The service provision by emergency vehicles is performed by private providers, who own and run several service centers equipped with emergency vehicles. The profit of a provider is considered to be proportional to transportation performance necessary for the demand satisfaction. In some national systems, the system administrator only defines some rules, under which a service provider is allowed to relocate his service centers [10].

The users' and providers' objectives are in a conflict, the user protecting rules comprise usual condition that the average or total value of disutility must not exceed a given limit. Following these rules, a considered provider, who performs reengineering, will change locations of his centers so that he maximizes the profit by capturing much demand under assumption that each demand is serviced from the nearest service center. We provided linear programming model of provider's reengineering of his part of emergency service system to maximize his profit under rules imposed by the system administrator [7]. In this paper, we focus on relation between the considered provider called leader and the other mutually cooperating providers, which represent a follower, which reacts on leader's decisions. The relation will be studied here as a form of Stackelberg's game [13].

2 Model of Follower's Reengineering

We introduce J as a finite set of all system users, where b_j denotes a volume of expected demand of user $j \in J$. Let I be a finite set of possible center locations. For each pair i, $j \in I \cup J$, the symbol d_{ij} denotes the integer distance between locations i and j, where the maximal relevant distance is denoted by m. The current emergency service center deployment is described by two disjoint sets of located centers $I_L \subset I$ and $I_F \subset I$, where I_L contains the centers of the leader and I_F is the set of p centers belonging to the other providers.

[1] Faculty of Management Science and Informatics, University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia, jaroslav.janacek@fri.uniza.s
[2] Faculty of Management Science and Informatics, University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia, marek.kvet@fri.uniza.s

The model below describes the problem, where positions of leader's service centers from I_L are fixed and the follower's service centers from I_F can be changed, for follower to improve his benefit. Due to some restrictions, the centers originally located at positions from I_F can be moved only to some free locations from I-I_L or they can be left unchanged. Let I_R denote the subset of possible center locations for the follower's centers. It holds that $I_F \subset I_R \subset I$-$I_L$.

The system administrator's rules are quantified by the following constants. The value F gives upper limit of the total transportation performance necessary for satisfaction of all users' demands (the total disutility perceived by system users). The value H is the maximal feasible distance between a user's location and the nearest service center.

We introduce coefficients a^s_{ij} for each pair i, j, where $i \in I_R \cup I_L$ and $j \in J$, where $a^s_{ij} = 1$ if and only if $d_{ij} \leq s$ and $a^s_{ij} = 0$ otherwise for $s = 0, 1, \ldots, m$-1.

We define cost coefficients for $i \in I_R$ and $j \in J$ so that $c_{ij} = 0$ if $d_{ij} \geq min\{d_{tj}: t \in I_L\}$ and $c_{ij} = b_j d_{ij}$ otherwise.

The last two auxiliary structures are denoted as $\{P_j\}$ and $\{R_j\}$, where $j \in J$. The first of them is a system of ordered lists, where list P_j consisting of $i \in I_R$ is ordered so that the following inequalities hold: $d_{P_j(1)j} \leq d_{P_j(2)j} \leq \cdots \leq d_{P_j(|I_R|)j}$. An element R_j of the second structure is an ordered list of subscripts from range $1, \ldots, |I_R|$, where $R_j(r)$ gives the minimal subscript, for which $d_{P_j(r)j} < d_{P_j(R_j(r))j}$ holds. Obviously $r+1 \leq R_j(r)$.

Now, we introduce series of decision variables, where binary variable y_i defined for each $i \in I_R$ takes the value of one, if a service center is to be located at i and it takes the value of zero otherwise.

To be able to express the total transportation performance value, we introduce zero-one auxiliary variables x_{js} for $j \in J$ and $s = 0, 1, \ldots, m$-1, where $x_{js} = 1$ if there is no located service center in the radius s from the user location j.

Finally, we introduce series of allocation variables $z_{ij} \in \{0, 1\}$ for $i \in I_R \cup I_L$ and $j \in J$, where $z_{ij} = 1$ if user demand located to j is serviced from center location i.

Using the above introduced structures and decision variables, we suggest the following model.

$$Maximize \sum_{j \in J} \sum_{i \in I_R} c_{ij} z_{ij} \tag{1}$$

$$Subject\ to \sum_{i \in I_R} y_i = p \tag{2}$$

$$\sum_{i \in I_R} a^H_{ij} y_i + \sum_{i \in I_L} a^H_{ij} \geq 1 \quad for\ j \in J \tag{3}$$

$$x_{js} + \sum_{i \in I_R} a^s_{ij} y_i + \sum_{i \in I_L} a^s_{ij} \geq 1 \quad for\ j \in J,\ s = 0, \ldots, m-1 \tag{4}$$

$$\sum_{j \in J} b_j \sum_{s=0}^{m-1} x_{js} \leq F \tag{5}$$

$$\sum_{i \in I_R \cup I_L} z_{ij} = 1 \quad for\ j \in J \tag{6}$$

$$z_{ij} \leq y_i \quad for\ j \in J,\ i \in I_R \tag{7}$$

$$1 - y_{P_j(k)} \geq \sum_{r=R_j(k)}^{|I_R|} z_{P_j(r)j} \quad for\ j \in J,\ k = 1, \ldots, |I_R| - 1 \tag{8}$$

$$y_i \in \{0, 1\} \quad for\ i \in I_R \tag{9}$$

$$x_{js} \in \{0, 1\} \quad for\ j \in J,\ s = 0, \ldots, m-1 \tag{10}$$

$$z_{ij} \in \{0, 1\} \quad for\ i \in I_R \cup I_L,\ j \in J \tag{11}$$

The objective function (1) expresses the volume of transportation performance allocated to the considered provider (provider's profit). If a user is nearer to a center of other providers, the contribution to the considered provider is zero. The misallocation of a user to a more distant center of the considered provider is prevented by constraints (8).

Constraint (2) preserves constant number of centers belonging to the considered provider under reengineering.

Constraints (3) ensure that any user j lies in the radius H from a located center, i.e. maximal distance between a user and the nearest center is less than or equal to the value H.

Constraints (4) give relation between located variables y_i and auxiliary variables x_{js} so that x_{js} equals to one, if no center is located in the radius s from the user's location j. Then, the expression $x_{j0}+ x_{j1}+\ldots+ x_{jm-1}$ gives the distance from the user j to the nearest service center regardless of its owner.

Constraint (5) makes use of the variables x_{js} and assures that the total transportation performance does not exceed the given value F.

Constraints (6) are commonly used allocation constraints, which assure that each user demand is allocated to exactly one center belonging either to the considered provider or to other providers.

Link-up constraints (7) give relation between allocation variables z_{ij} and the location variables y_i, which model the decisions on locating service centers operated by considered provider.

Constraints (8) were developed to prevent the maximization process from allocating a demand to a more distant service center than the nearest one. The constraint formulated for location $P_j(k)$ and user j forbids allocation of user's j demand to every service center $P_j(r)$, which is more distant from the location j than the center location $P_j(k)$.

In remainder of this paper, the optimal set of the follower's center locations obtained by solving the problem (1) – (11) for a set I_L of fixed centers and for a given matrix $\boldsymbol{c}$ of cost coefficients will be denoted as $I^c(I_L)$.

3 Stackelberg's Game of Service Providers

We shall study the problem of a leading provider, which starts with a rearrangement of the set I_L of his service center locations. The leader's action is followed by the follower's changes of follower's service center locations to optimize his profit subject to the current deployment of the leader's service center. Let us express leader's profit from a given system user j under assumption that leader's and follower's sets of service centers are known. Let I_L and I_F denote the current leader's and follower's sets of centers respectively. As a user's demand is assigned to the nearest service center, the leader's profit $C_j(I_L, I_F)$ from the user j can be defined by (12).

$$\begin{aligned} &\textit{If } \min\{d_{ij} : i \in I_L\} > \min\{d_{ij} : i \in I_F\} \textit{ then } C_j(I_L,\ I_F) = 0 \\ &\textit{else } C_j(I_L,\ I_F) = b_j \min\{d_{ij} : i \in I_L\} \end{aligned} \tag{12}$$

The total leader's profit can be expressed by (13).

$$\sum_{j \in J} C_j(I_L, I_F) \tag{13}$$

The leader's problem of this Stackelberg's game can be described as a task of determination of such set I_L that the profit (13) reaches the maximal value taking into account the subsequent follower's reaction. It is assumed that the leader is not allowed to place any his center at locations currently occupied by the follower. The leader's problem with p centers can be described as follows.

$$\max\{\sum_{j \in J} C_j(I_L, I^c(I_L)) : I_L \subset I - I_F,\ |I_L| = p\} \tag{14}$$

Unfortunately, the set $I^c(I_L)$ is a result of an optimization process performed by an IP-solver with the problem (1)-(11) and thus the set cannot be expressed analytically. This is why; the problem (14) cannot be represented by one mathematical programming model and solved by standard tools of mathematical programming. The existing approaches to this problem made use of heuristics and their authors tried to obtain a good solution of (14) by a searching process, where a heuristics for $I^c(I_L)$ determination was embedded into heuristic process searching for I_L [2], [9], [11].

We suggest a different approach, which makes use of the exact solving technique for $I^c(I_L)$ mentioned in Section 2. The approach is based on an estimation of the follower's future activities.

4 Approximate Solution of Leader's Problem

Having assumed that the follower's set I_F stays unchanged, we can solve the leader's problem to optimality by the above-mentioned process, which produces the optimal set of leader's center locations [7]. The process uses profit coefficients c_{ij} for $i \in I\text{-}I_F$ and $j \in J$ so that $c_{ij} = 0$ if $d_{ij} \geq min\{d_{tj}: t \in I_F\}$ and $c_{ij} = b_j d_{ij}$ otherwise. This way of leader's problem solution would be appropriate only under the assumption that the follower will not react on the leader's move (change of I_L). The suggested approach tries to anticipate follower's reaction on the leader's move such that it adjusts the profit coefficients taking in the account the situation which can appear after the leader's move.

Our approach to the leader's problem is based on the modified profit coefficients, which reflect the possibility that the follower places his center so that it will be nearer to the user than a leader's center. We come from the hypothesis that the original contribution $b_j d_{ij}$ of a user j to the leader's center located at i will be diminished proportionally to the number of possible center locations, which distance from the user location j is less than or equal to the distance between locations i and j. Let us denote the number of possible locations j and i by the symbol n_{ij}. Then, the modified coefficients $\underline{c}_{ij}$ will be determined according to (15).

$$\begin{aligned} &If \;\; d_{ij} \geq \min\{d_{tj} : t \in I_F\} \; then \; \underline{c}_{ij} = 0 \\ &else \; \underline{c}_{ij} = b_j d_{ij} / (n_{ij})^{\alpha} \end{aligned} \tag{15}$$

The non-negative exponent α in (15) is a parameter of this strategy. Obviously, an increase of the parameter leads to diminishing of the expected profit. The leader's solution, i.e. his new service center locations can be obtained as the result of the solving process applied on the model (1)-(14), where coefficients $\underline{c}_{ij}$ are employed instead of c_{ij} and I_F is used instead of I_L, will be denoted as $I^{\underline{c}}(I_F)$. Applying the leader's solution into the Stackelberg's game, $I^{\underline{c}}(I_F)$ corresponds to the result of the leader's move and this result can be used in the optimization process to obtain the result $I^c(I^{\underline{c}}(I_F))$ of the follower's move. Then, the resulting profit of the leader after the both steps of the game are performed is expressed by (16).

$$\sum_{j \in J} C_j(I^c(I_F), I^c(I^c(I_F))) \tag{16}$$

5 Numerical Experiments

To study presented approximate approach to Stackelberg's game, we performed series of numerical experiments, in which the optimization software FICO Xpress 8.3 (64-bit, release 2017) 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. We used benchmarks derived from real emergency health care system of the Slovak Republic, which we partitioned into six self-governing regions, i.e. Bratislava (BA), Banská Bystrica (BB), Nitra (NR), Trenčín (TN), Trnava (TT) and Žilina (ZA). In each benchmark, all cities and villages were taken into account. The corresponding number b_j of the dwelling place inhabitants was rounded to hundreds. The set of dwelling places represents both the set J of users' locations and the set I of possible center locations as well. The numbers of dwelling places in the individual self- governing regions are reported in Table 1, where the associated column is denoted by $|I|$. The set of currently operated centers in a self-governing region is denoted by $I_1 \subset I$ and the number of located service centers in the individual self-governing regions is reported in the column $|I_1|$. The network distance from a user to the nearest located center was considered as the user´s disutility.

Table 1: Size of used benchmarks.

Region	$\|I\|$	$\|I_1\|$	$\|I_L\|$	$\|I_F\|$
BA	87	14	8	6
BB	515	36	17	19
NR	350	27	16	11
TN	276	21	10	11
TT	249	18	11	7
ZA	315	29	15	14

Each benchmark was completed by random partitioning of the set I_I into two subsets I_L and I_F of leader's and follower's centers respectively so that approximately half of the centers belong to the leader. The exact cardinalities of the sets I_L and I_F are presented in Table 1 in the columns denoted by $|I_L|$ and $|I_F|$.

The current profit of the leader and follower is plotted at Table 2 for each benchmark in the columns SP_L and SP_F respectively. There is also shown how the leader's profit drops if he resigns to his move and let move the follower only. The resulting profits are given in columns FP_L and FP_F of Table 2.

The substantial part of the numerical experiments is devoted to exploration of the suggested approach to the leader's problem. First of all, our attention was focused on proper setting of the parameter α. Based on our preliminary experiments, we established range of the parameter as interval [0, 4] and performed enumeration of the expression (14), which gives the resulting leader's profit subject to given value of the parameter α. The resulting values are plotted in Table 3.

Table 2: Profit shares of the leader and follower at the initial positions and after follower's only move.

Region	Current state		Only follower's move	
	SP_L	SP_F	FP_L	FP_F
BA	8965	12877	8282	13290
BB	17594	14882	16518	15956
NR	24822	14009	19648	19160
TN	15151	11532	10962	15628
TT	19927	11655	16782	14735
ZA	15437	16518	12063	19890

Table 3: Resulting leader's profit, when leader's strategy given by parameter α is used.

Region\α:	0	0.125	0.25	0.5	0.75	1	2	4
BA	9358	9358	9358	8787	8787	9344	11042	6843
BB	17328	17282	17282	17282	16274	16274	16274	14828
NR	23312	23312	23121	24314	24728	24181	23958	22039
TN	11854	11854	12212	12212	12212	12212	12212	11417
TT	19560	19441	19100	18812	18812	19240	18027	19192
ZA	14420	14375	14696	14696	15239	14446	13409	12715
AVG	15972	15937	15962	16017	16009	15950	15820	14506

6 Conclusions

The paper deals with an approach to emergency service system reengineering, where the service providers play so-called Stackelberg's game, in which they compete for profit following transportation performance necessary for servicing their clusters of users. The clusters are formed so that a customer is assigned to the nearest located service center and the user's contribution to profit of the center owner is proportional to product of the distance between user's and center locations and the user's demand. The game is played so that the considered provider – leader performs his move first. The move consists in changing positions of his centers to improve his profit. But, this leader's move is followed by follower's reaction, which is performed with knowledge of the leader's changes. This reaction may considerably reduce the resulting leader's profit. This way, leader's strategy has to anticipate the follower's reaction on the leader's move.

We have suggested leader's strategy based on adjustment of profit coefficients, which reflect the possibility that the follower places his center so that it will be nearer to the user than a leader's center. As the suggested strategy depends on setting of the parameter α, we have performed a research of its suitable setting and presented the results.

The presented approach may serve as a useful tool for research of providers' behavior in reengineering processes, in which different groups of providers compete for the profit under system administrator supervision. Future research may be aimed at fair solution of the reengineering problem.

Acknowledgements

This work was supported by the research grants VEGA 1/0342/18 "Optimal dimensioning of service systems", VEGA 1/0463/16 "Economically efficient charging infrastructure deployment for electric vehicles in smart cities and communities" and APVV-15-0179 "Reliability of emergency systems on infrastructure with uncertain functionality of critical elements".

References

[1] Brotcorne, L. Laporte, G. & Semet, F. (2003). Ambulance location and relocation models. *European Journal of Operational Research* 147, 451–463.
[2] Consoli. S. & Moreno-Pérez. J. A. Mladenovic, N. & Darby-Dowman, K. (2010). Discrete article swarm optimization for the minimum labeling spanning tree problem. *Nat. Comput*. 9, 29-46.
[3] Doerner, K.F., et al. (2005). Heuristic Solution of an Extended Double-Coverage Ambulance Location Problem for Austria. *Central European Journal of operations research*, Vol. 13, No 4, 325-340.
[4] Elloumi, S. Labbé, M. & Pochet, Y. (2004). A new formulation and resolution method for the p-center problem. *INFORMS Journal on Computing* 16, 84-94.
[5] García, S. Labbé, M. & Marín, A. (2011). Solving large p-median problems with a radius formulation. *INFORMS Journal on Computing*, Vol. 23, No 4, 546-556.
[6] Janáček, J. & Kvet, M. (2013). Public service system design with disutility relevance estimation. In *Proceedings of the Mathematical Methods in Economics 2013*, Jihlava, Czech Republic, 332-337.
[7] Janáček, J. & Kvet, M. (2018). Reengineering of the Emergency Service System from the Point of Service Provider. In *Proceedings of the ICORES 2018: 7th International Conference on Operations Research and Enterprise Systems*, Madeira, Portugal, 224-231. accessible at: http://www.scitepress.org/ProceedingsDetails.aspx?ID=OFC53CqGGsA=&t=1
[8] Jánošíková, Ľ. & Žarnay, M. (2014). Location of emergency stations as the capacitated p-median problem. In *International scientific conference: Quantitative Methods in Economics-Multiple Criteria Decision Making XVII*, Virt, Slovakia.
[9] Karakitsu, A. (2010). *Modeling Discrete Competitive Facility Location*.Heidelberg, Springer
[10] Kvet, M. & Janáček, J. (2016). Radiálny prístup na zlepšenie existujúceho záchranného systému. In *Optimalizační úlohy v dopravních a logistických systémech a SW podpora rozhodování v inteligentních dopravních systémech*, Praha, Czech Republic, 11-25.
[11] Rodríguez C. C. Moreno-Pére, J. A. & Santos-Penate, D. R. (2011). Particle swarm optimization with two swarms for the discrete (r|p)-centroid problem. In *Computer Aided Systems Theory* – EUROCAST 2011, Lecture Notes in Computer Science, Berlin, Springer, 432-439.
[12] Sayah, D. & Irnich, S. (2016). A new compact formulation for the discrete p-dispersion problem. *European Journal of Operational Research*, Vol. 256, No 1, 62-67.
[13] von Stackelberg, H. (1952). *The Theory of the Market Economy*, William Hodge, London.

Semantic Model of Management in Student Projects

Tereza Jedlanová[1], Jan Bartoška[2], Klára Vyskočilová[3]

Abstract. The article describes the use of semantic networks and analytical network analysis (ANP) methods for quantification of the "soft" structure of a project. Design and use of the semantic model is shown on selected student projects. The semantic project networks are derived based on the Work Breakdown Structure and the RACI matrix. Their subsequent quantification using the ANP method creates the basis for a team communication in organization. Although the prioritization of project roles in communication or project documentation in project management is key to the success of the project, a unique approach to quantify the soft structure of the project has not been introduced yet. Semantic networks can be used to manage projects to illustrate and quantify links between internal and external objects of the project environment -- project goals, project outputs and project documents. The paper suggests a new approach to identifying and quantitating project roles in the project communication. The showed semantic model and its quantification can be used to construct a multi-criteria model for decision support in project management. This may be applied not only in student organizations but also in commercial practice.

Keywords: Project Management, Student Projects, Semantic Networks, Stakeholder Management, Analytic Network Process, Human Resource Project Management, Multi-criteria Decision Making.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

So far in project management no analytical tool has been implemented for a comprehensive analysis of the management structure, with the possibility of quantification of elements and links for project roles, project documentation, project constraints or project outputs. The proposed tool to perform this task in the form of semantic networks and ANP (Analytic Network Process) is suggested by [1] and [2]. Its use makes it possible to obtain quantitative feedback on the significance of elements in the project structure and its management - especially in the case of a project team.

Project management is a collection of many professional disciplines and skills when, besides traditional quantitative approach, qualitative approach which is the content of international standards is gaining on importance. International standards of project management, i.e. PMBOK®Guide [9], PRINCE2 [7] and ISO 21500:2012 [4] describe a common practice to set organizational, instrumentation, process and knowledge management aspects of a project, giving rise to the so-called "soft" structure of the project, i.e. a partially ordered set of documents, roles, workflows and tools with variable influence of internal and external environment of the project. Although the prioritization of project roles in communication and project documentation in project management is a key for the success of the project, a unique approach to quantifying the "soft" structure of the project has not yet been introduced. "Soft" structures of the project are usually just displayed in the form of knowledge maps or semantic networks. For potential quantification it is appropriate to focus particularly on the use of semantic networks.

Semantic networks are conventionally utilized in ICT, especially in developing and maintaining software and web applications, e.g., Kim et al. [5] uses a semantic model for viewing and extracting structure tags for creating and managing web sites, or Zhu and Li [18] focus on developing a semantic approach in IT for the resolution of an application and contextual level. The use of semantic networks can lead to the improvement of the systems studied, such an example is a case study in the field of IT by Shing-Han, Shi-Ming, David and Jui-Chang [15]. Semantic networks are also used for expressing the relationships between users of information systems and content

[1]Department of Systems Engineering FEM CULS, Kamýcká 129, Prague, jedlanova@pef.czu.cz.
[2]Department of Systems Engineering FEM CULS, Kamýcká 129, Prague, bartoska@pef.czu.cz.
[3]Department of Systems Engineering FEM CULS, Kamýcká 129, Prague, vyskoc.klara@gmail.com.

web servers, where significance quantification of elements is subsequently conducted with the help of the PageRank method [8]. In a similar manner, but using the ANP method (Analytic Network Process) [13], it is possible to quantify the elements of the "soft" project structure, i.e. e.g. project roles, project documentation and project constraints. The use of semantic networks in project management is not yet prevalent. For the environment of IT project models, Schatten [14] proposes semantic social networks for the extraction and transfer of knowledge within the project team and stakeholders of the project. Whereas Williams [17] uses a semantic network for the display and analysis of success factors in the structure of mutual expectations and limitations of the project stakeholders. While El-Gohary, Osman and El-Diraby [3] for example indicate that insufficient or inadequate stakeholders' involvement in the management of projects, programs and portfolios is the main reason for failure, and they further propose the semantic model as a tool to visualize and manage relationships and knowledge towards stakeholders in a multi-project environment. Following the example of their use in IT, it is therefore appropriate to continue developing semantic models to meet the needs of project management, i.e. a management structure of roles, documentation, and knowledge constraints in a project.

The aim of the article is to use a semantic model for the management of projects in the student project environment and to quantify the significance of project roles. The benefit should be a holistic approach to identifying and quantifying project roles as elements in the structure of the project environment.

2 Materials and methods

2.1 Project Management Institute – PMBOK® Guide

Project Management Institute (PMI) ranks among the largest non-profit associations in the world, dealing with project management. The association was founded in the United States in 1969. Since the 80s of the 20th century, in a broad discussion among experts and professional public, it has been developing a standard entitled "A Guide to the Project Management Body of Knowledge" (PMBOK®Guide).

The PMBOK®Guide [9] is process-oriented and in its knowledge areas it presents the set of qualitative (e.g. team management, team motivation, etc.) or quantitative approaches (e.g. CPM, EVM, etc.) necessary for project management. Knowledge areas of the PMBOK®Guide standard [9] represent the substantive content of the project lifecycle processes. The standard thus dissolves project management into atomic transformations where the entry is secured, necessary tasks (expert, professional, managerial, etc.) are performed and output is created. Outputs from one process become inputs of the second.

2.2 PRINCE2 – Projects in Controlled Environments

International project management methodology PRINCE2 originated at Simpact Systems Ltd. in 1975 as a methodology of PROMPT projects management. In 1979, the UK government has accepted PROMPT II methodology as the official methodology for managing IT projects in public administration. In 1989, the methodology was expanded and published under the new name PRINCE - "Projects in Controlled Environments". Using the topics [7], the methodology indirectly defines its own knowledge areas, i.e. the areas to be managed across processes, with the help of specific expertise and with a specific goal.

The topics of PRINCE2 methodology are as follows [7]: Business Case; Organization; Quality; Plans; Risk; Change; Progress. The seven processes which PRINCE2 methodology provides describe the life cycle of a project from the position of roles and responsibilities. All seven PRINCE2 processes can be continuously connected. PRINCE2 methodology processes [7] are based on the interaction of project roles and stakeholders in the organizational structure of the project. Project roles and roles of stakeholders in PRINCE2 are as follows [7]: Steering Committee; Sponsor; Senior User; Senior Supplier; Project Manager; Team Manager; Change Authority; Project Support; Project Assurance.

2.3 Semantic Model

A semantic model consists of a semantic network that Mařík et al. [6] define as "natural graph representation", where each node of a graph corresponds to a specific object and each edge corresponds to a binary relation. Mařík et al. [6] further state that "semantic networks can conveniently express the relations of set inclusion and membership in a set, and unique as well as general terms can be represented there". Semantic networks emerged at the end of the 60s of the 20th century. The term "semantic network" was for the first time used by Quillian [10] in his dissertation on the representation of English words. According to Sow [16], semantic networks are used for their ability to provide an easy-to-use system of information representation. The semantic networks are suitable for displaying and expressing vast information resources, management structures and processes or other areas, e.g. [11].

2.4 Analytic Network Process (ANP)

The ANP (Analytic Network Process) method is the generalization of the AHP (Analytical Hierarchy Process) method. The ANP model reflects and explores the increasing complexity of network structure, where the network is made up of different groups of elements. Each group (cluster) comprises a homogeneous set of elements. Linkages can exist among clusters as well as among the elements. In the ANP model the hierarchy is removed [13].

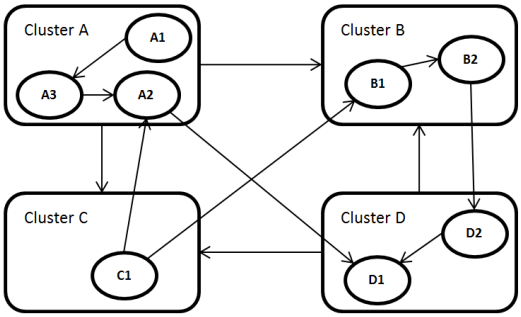


Figure 1 The **s**tructure of elements and clusters in ANP model

The benefit of the ANP method is the ability to express different preferences of links between elements and clusters. To express preferences the method of pair comparison is used. Preferences always occur precisely in assessing the importance of the two elements in terms of the element which refers to them - there rises a question "which of the elements is more important, and by how much". The resulting values for the sub-clusters are then combined into a super matrix where the normalisation of columns is performed [12] or [13].

$$W = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_N \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \dots & C_N \end{array} \\ \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \dots & W_{nn} \end{bmatrix} \end{array} \tag{1}$$

Where each block of the super matrix consists of:

$$W_{ij} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} \tag{2}$$

Under condition:

$$\sum_{i}^{n} w_{ij} = 1, j \in \langle 1, n \rangle \tag{3}$$

For a weighted super matrix (1) for which the relation (3) is valid, a calculation to obtain limit weights of the elements can be performed. The calculation is performed by squaring the weighted super matrix to a sufficiently large number. Since the super matrix has NxN size, the squaring is always feasible in a trivial manner (matrix multiplication).The result is the approximation of the weighted matrix to the limit matrix. Limit scales can be found in any column of the super matrix. The limit weight of each element expresses the strength of the effect on the overall structure of elements, i.e. it answers the question of how strongly an element affects the other elements [12] or [13].

2.5 Management in Student Projects

For the purposes of designing a semantic project management network for student projects, the Work Breakdown Structures, the RACI matrix, and the organizational structure organogram with roles in student projects are used. According to [9] and [7], the Work Breakdown Structure (WBS) is a tool in the form of a tree chart in which the target (root of the tree) is decomposed into partial project outputs, work packages and, consequently, atomic tasks. The WBS is complemented by the creation of an organizational structure, which in the case of student projects has a form as illustrated in Figure 2.

After creating the Work Breakdown Structure and organizational structure chart, it is possible to define a RACI matrix for managing human resources in a team. According to [9] and [7], RACI matrices are incidental in nature - in columns defined roles are assigned partial activities in a row. The incidence of the row and column represents the type of responsibility:

- R - Responsible - person working on an activity,
- A - Accountable - person accountable for a fulfilled task,
- C - Consulted - person providing consultancy, giving answers when asked,
- I - Informed - person receiving information in order to know what is happening and not expected to react.

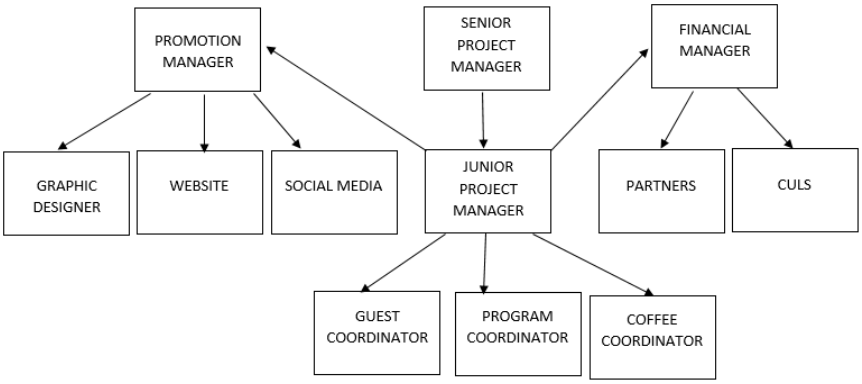


Figure 2 Organizational structure of student projects

3 Results and Discussion

3.1 Creating a Semantic Model for Student Projects

Based on [1] and [2], two student projects were selected to create the semantic model of project management and the ANP method: "*Ludo*" and "*Wine festival*". The creation of a semantic model is based on a progressive use of the Work Breakdown Structure (WBS) [9], where the core activities of the project are located, and the RACI matrix [7] which identifies dependence on project roles and project activities that bind their results on project outputs and documentation, or its course on project constraints.

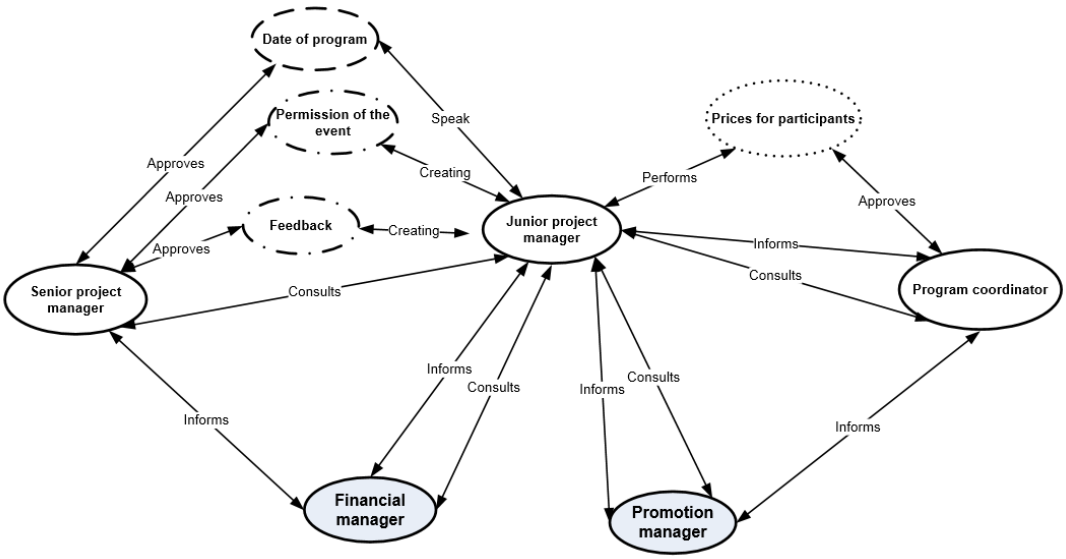


Figure 3 The part of semantic network of the project "*Wine Festival*"

In the case of student projects clusters were also proposed for the creation of semantic models:

- *Role* - cluster of elements representing partial roles in the project team;
- *Outputs* - clusters of elements that represent planned outputs of the project in the form of expected products, services or returns (benefits);

- *Documentation* - a cluster of elements that present project documentation for project planning and management;
- *Restrictions* - cluster of elements that represent project constraints, dependencies or partial threats.

The clusters that define the structure of the projects are made up of partial elements. The nature of the elements is always given by the cluster they belong to. Names of the elements, such as Documents or Outputs, differ for each project. In addition, for the purposes of creating semantic models for student projects, the following linkages were defined:

- R (responsibility) – the linkage goes from a given role and is of the "conjugate - perform - create" type;
- A (accountability) – the link goes from the role and is of the "approve" type;
- C (consulted) – the link goes from role R to role C and is of the "consult" type;
- I (informed) – the link goes from role R to role I and is of the "inform" type.

The linkages express the possible interaction among the elements, i.e. project roles, documentation, outputs or project constraints. The resulting topology of the semantic model differs between the selected student projects - the arrangement of the linkages corresponds to the process of project management and planning and implementation. For illustration, Figure 3 presents a sub-network of a semantic model for the "Wine Festival" student project. Thanks to the procedure described here, i.e. the progressive use of WBS and RACI matrix tools, cluster and linkage definition, it was possible to clearly identify the nature and task of project roles in the structure and management of the project.

3.2 Quantification of Project Roles in Project Structure

The semantic model quantification of the project structure can be performed by the ANP method. The advantage of this method is the possibility of bias preferences among elements. Creating the calculation of the ANP model (Figure 2) can be performed e.g. in a software tool MS Excel. By calculating a super matrix can be obtained weights of the project roles. The resulting weights of the individual elements reflect the influence of the element on the structure of the whole model even within individual clusters:

"*Wine festival*" Project		"*Ludo*" Project	
Weight	**Role**	**Weight**	**Role**
0.199052	Senior project manager	0.171429	Senior project manager
0.170616	Junior project manager	0.171429	Junior project manager
0.14218	Financial manager	0.114286	Financial manager
0.14218	Promotion manager	0.114286	Promotion manager
0.113744	Coordinator of participants	0.142857	Coordinator of participants
0.113744	Program coordinator	0.142857	Program coordinator
0.118483	Food coordinator	0.142857	Food coordinator

Table 1 Quantification of sub-roles in the project team

From the results achieved in both student projects it is clear that the highest weight and thus the highest importance is related to responsibility. In both projects and project teams, it is the Senior Project Manager and Junior Project Manager who have the highest responsibility for the project. Both roles have the highest weight in the project management structure. The significance of other elements in the cluster is gradually decreasing according to the real responsibility in the project - the decrease in the projects varies in relation to their nature and difficulty. The semantic model and quantification of roles demonstrate the realistic organization of the project team and the project management structure.

4 Conclusion

The semantic project management model can be used to display and quantify elements of a project structure. The weighting of elements contributes to the clarification of the significance of project roles. The results can be used for control (in terms of frequency) and reward (based on position) of team members, or to ensure communication.

The article describes how to identify and quantify project roles in a project structure, in the context of planned outputs, project documentation, and project constraints. The project management tools, namely the Work Breakdown Structure to view key project activities, organizational structure of the project (interception of all project roles) and, last but not least, the RACI matrix, which was expanded with possible outcomes from individual activities, were used. The research was conducted on two student projects bearing the same organizational structure and comparable activities during the project. Differences were only in size and focus of the projects. The achieved

results have highlighted the difference in importance of roles for each project – comparable projects differ in the interaction among roles in the project structure.

References

[1] Bartoška, J. (2016a). *A sematic model for the management of projects*. Habilitation work (thesis). DSE FEM CULS Prague.

[2] Bartoška, J. (2016b). Semantic Model of Organizational Project Structure. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics*. Liberec: Technical University. ISBN 978-80-7494-296-9.

[3] El-Gohary, N. M., Osman, H. and El-Diraby, T.E. (2016). Stakeholder management for public private partnerships. *International Journal of Project Management, 24*, 595-604.

[4] *ISO 21500:2012 Guidance on project management*. Geneva: International Organization for Standardization, 2012.

[5] Kim, H. L., Decker, S. and Breslin, J. G. (2010). Representing and sharing folksonomies with semantics. *Journal of Information Science 36*, 57–72.

[6] Mařík, V. et. al. (1993). *Artificial Intelligence I.* Prague: Academia. ISBN 80-200-0496-3.

[7] Murray, A. et. al. (2013). Managing Successful Projects with PRINCE2®. Fifth edition. London: AXELOS Limited. ISBN 978-0113310593.

[8] Page, L., Brin, S., Motwani, R. and Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report. Stanford: Stanford InfoLab.

[9] Project Management Institute (2013). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*. Fifth Edition. Newtown Square (PA): Project Management Institute Inc., 14 Campus Boulevard. ISBN 978-1-935589-67-9.

[10] Quillian, R. (1968). *A recognition procedure for transformational grammars.* Doctoral dissertation. Cambridge (MA): Massachusetts Institute of Technology.

[11] Rydval, J., Bartoška, J., Brožová, H. (2014). Semantic Network in Information Processing for the Pork Market. *AGRIS on-line Papers in Economics and Informatics 6*, 59-67.

[12] Saaty, T. L. and Vargas, L. G. (2006). *Decision Making with the Analytic Network Process: Economic, Political, Social and Technological Application with Benefits, Opportunities, Costs and Risks*. New York: Springer.

[13] Saaty, T. L. (1996). *Decision Making with Dependence and Feedback: The Analytic Network Process*. Pittsburgh: RWS Publisher.

[14] Schatten, M. (2013). Knowledge management in semantic social networks. *Computational and Mathematical Organization Theory 19*, 538-568.

[15] Shing-han, L., Shi-ming, H., David and C. Y., Jui-chang, S. (2013). Semantic-based transaction model for web service. *Information Systems Frontiers 15*, 249-268.

[16] Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations.* Pacific Grove (CA): Brooks/Cole Publishing Co.

[17] Williams, T. (2015). Identifying Success Factors in Construction Projects: A Case study. *Project Management Journal 47*, 97-112.

[18] Zhu, Y. and Li, X. (2007). Representations of semantic mappings: A step towards a dichotomy of application semantics and contextual semantics. *International Journal of Project Management 25*, 121-127.

Multi–Objective Optimization Problems with Random Elements; Survey of Approaches

Vlasta Kaňková [1]

Abstract.

Many economic and financial situations depend simultaneously on a random element and a decision parameter. Mostly, it is possible to influence the above mentioned situation only by an optimization model depending on a probability measure. This optimization problem can be static (one–stage), dynamic with finite or infinite horizon, single–objective or multi–objective. We focus on one–stage multi–objective problems corresponding to applications those are suitable to evaluate simultaneously by a few objectives. The aim of the contribution is to give a survey of different approaches (as they are known from the literature) of the above mentioned applications. To this end we start with well–known mean–risk model and continue with other known approaches. Moreover, we try to complete every model by a suitable application. Except an analysis of a choice of the objective functions type we try to discuss suitable constraints set with respect to the problem base, possible investigation and relaxation. At the end we mention properties of the problem in the case when the theoretical"underlying" probability measure is replaced by its "deterministic" or "stochastic" estimate.

Keywords: Multi–objective optimization problems, random element, mean–risk model, deterministic approach, stochastic multi–objective problems, constraints set, relaxation

JEL classification: C44
AMS classification: 90C15, 90C29

1 Introduction

Multi–objective optimization problems with a random element correspond to many economic situations in which an economic process is influenced by a random factor say ξ, a decision parameter say x and it is suitable to evaluate it by a few objectives. To recall an exact mathematical definition of the optimization problem depending on a random factor, let $(\Omega, \mathcal{S}, P)$ be a probability space; $\xi := \xi(\omega) = (\xi_1(\omega), \ldots, \xi_s(\omega))$ s–dimensional random vector defined on $(\Omega, \mathcal{S}, P)$; $F(:= F(z),\ z \in R^s)$, P_F and Z_F denote the distribution function, the probability measure and the support corresponding to ξ; $x = (x_1, \ldots, x_n) \in R^n$. Let, moreover, $g_i := g_i(x, z)$, $i = 1, \ldots, l$, $l \geq 1$, $\quad g_j^* := g_j^*(x, z)$, $j = 1, \ldots, l'$, $\quad l' \geq 1$ be real–valued (say continuous) functions defined on $R^n \times R^s$, $\quad X \subset R^n$ be a nonempty set.

The above mentioned rather general multi–objective problem with a random element (in static setting) can be introduced in the following form:

$$\begin{array}{ll} \text{Find} & \min g_i(x, \xi),\ i = 1, \ldots, l \\ \text{subject to} & g_j^*(x, \xi) \leq 0,\ j = 1, \ldots, l', \quad x \in X. \end{array} \tag{1}$$

(1) is (generally) a non linear multi–objective (practically "deterministic") programming problem everywhere when the decision x can depend on the random element ξ, it means when the realization ξ is known in the time of the problem solution. However, it is known that such "nice" situation happen very seldom.

[1] The 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

On the other hand rather often it is possible to assume that the probability measure P_F is known or at least it can be estimated. Consequently, (in such case) a question arises how to determine a suitable problem. Of course, it is possible to assume that the problem will be again multi–objective and that it will be depending on a probability measure. To explain this situation, let us start with very simple and well known one–objective portfolio problem (1) in which $l = l' = 1$, the functions g_1^* is not depending on the random element ξ. The problem (with random element) is the following:

$$\text{Find} \quad \max \sum_{k=1}^{n} \xi_k x_k \quad \text{s.t.} \quad \sum_{k=1}^{n} x_k \leq 1, \quad x_k \geq 0, \quad k = 1, \ldots, n, \quad n = s, \tag{2}$$

(x_k is a fraction of the unit wealth invested in the asset k; ξ_k return of the asset k). Evidently, if it is necessary to determine x_k without knowledge the realization ξ_k, $k = 1, \ldots, n$, then it is (very often) reasonable to set to (2) two–objectives "deterministic" optimization problem:

$$\text{Find} \quad \max \sum_{k=1}^{n} \mu_k x_k; \qquad \max[-\sum_{k=1}^{n}\sum_{j=1}^{n} x_k c_{k,j} x_j], \quad \text{s.t.} \quad \sum_{k=1}^{n} x_k \leq 1, \quad x_k \geq 0, \quad k = 1, \ldots, n, \tag{3}$$

in which $\mu_k = \mathsf{E}_F \xi_k, \quad c_{k,j} = \mathsf{E}_F(\xi_k - \mu_k)(\xi_j - \mu_j)$. (Symbol E_F denotes the operator of mathematical expectation corresponding to the distribution function F; of course we suppose that the corresponding final mathematical expectation exists).

Remark 1. The "underlying" problem with random element (2) is single –objective with deterministic constraints, the corresponding problem (3) depending on the probability measure is two–objectives, where one is in a form of mathematical expectation and the other is given by the second moment; constraints set is deterministic.

To find x that maximize simultaneously both objectives is mostly impossible. Markowitz in [13] suggested to replace the problem (3) by one–objective problem:

$$\text{Find} \quad \min[-\sum_{k=1}^{n} \mu_k x_k + K \sum_{k=1}^{n}\sum_{j=1}^{n} x_k c_{k,j} x_j] \quad \text{s.t.} \quad \sum_{k=1}^{n} x_k \leq 1, \quad x_k \geq 0, \quad k = 1, \ldots, n, \, K \geq 0. \tag{4}$$

The Markowitz approach to the portfolio problem (3) have started the general approach to multi–objective stochastic optimization problems, known (from the literature) as scalarizing. (This approach is well known also from the deterministic multi–objective problems theory.)

2 Scalarizing

To explain this approach we start with more general multi–objective problem with random elements and suppose: the decision vector x has to be determined without knowledge of the random element realization ξ. If it is reasonable to determine the decision with respect to the mathematical expectation of the objectives; the constraints set can be included in "deterministic" constraints depending on P_F, then the corresponding multi–objective stochastic problem can be introduced in the form:

$$\text{Find} \quad \min \mathsf{E}_F g_i(x, \xi), \; i = 1, \ldots, l \quad \text{s.t.} \quad x \in X_F. \tag{5}$$

Remark 2.

a. We assume that the final mathematical expectation $\mathsf{E}_F g_i(x, \xi)$, $\mathsf{E}_F g_j^*(x, \xi)$, $i1, \ldots, l$, $j = 1, \ldots, l'$ exist for all $x \in X$.

b. The mathematical expectation in (5) can be replaced by another functional (see, e.g., the relation (4)); we denote it generally by a symbol $\mathcal{F}$.

c. Considering (1) with g_j^* omitted, the following problem in [4] is considered:

$$\text{Find} \quad \min_{x,\,u}(u_1, \ldots, u_l) \quad \text{s.t.} \quad P_F\{g_i(x, \xi) \leq u_i\} \geq \beta_i, \quad \beta_i \in \langle 0, 1\rangle, \quad i = 1, \ldots, l, \quad x \in X.$$

Simultaneously, a comparison with problems mean, mean – variance, mean – standard deviation is introduced in [4].

A few types $X_F \subset R^n$ are known from the stochastic programming literature, they are "deterministic" ($X_F = X$); given by mathematical expectation $\mathsf{E}_F g_j^*$, $j = 1, \ldots, l'$; determined by probability constraints (for more details see see, e.g., [1]); recently rather often stochastic dominance constraints appear. The results obtained for deterministic multi–objective problems enable to apply these cases.

2.1 Deterministic Problems

To recall the suitable results obtained for deterministic problems, let $f_i(x)$, $i = 1, \ldots, l$ be real–valued functions defined on R^n; $\mathcal{K} \subset R^n$ be a nonempty set. The multi–objective problem can be defined by:

$$\text{Find} \quad \min f_i(x),\ i = 1, \ldots, l \quad \text{subject to} \quad x \in \mathcal{K}. \tag{6}$$

Definition 1. The vector x^* is an efficient solution of the problem (6) if and only if there exists no $x \in \mathcal{K}$ such that $f_i(x) \leq f_i(x^*)$ for $i = 1, \ldots, l$ and such that for at least one i_0 one has $f_{i_0}(x) < f_{i_0}(x^*)$.

Definition 2. The vector x^* is a properly efficient solution of the multi–objective optimization problem (6) if and only if it is efficient and if there exists a scalar $M > 0$ such that for each i and each $x \in \mathcal{K}$ satisfying $f_i(x) < f_i(x^*)$ there exists at least one j such that $f_j(x^*) < f_j(x)$ and

$$\frac{f_i(x^*) - f_i(x)}{f_j(x) - f_j(x^*)} \leq M. \tag{7}$$

Proposition 1. [7] Let $\mathcal{K} \subset R^n$ be a nonempty convex set and let $f_i(x)$, $i = 1, \ldots, l$ be convex functions on $\mathcal{K}$. Then $x^0 \in \mathcal{K}$ is a properly efficient solution of the problem (6) if and only if x^0 is optimal in

$$\min_{x \in \mathcal{K}} \sum_{i=1}^{l} \lambda_i f_i(x) \quad \text{for some} \quad \lambda_1, \ldots, \lambda_l > 0, \quad \sum_{i=1}^{l} \lambda_i = 1.$$

A relationship between efficient and properly efficient points is introduced, e.g., in [6] or [7]. We summarize it in the following Remark.

Remark 3. Let $f(x) = (f_1(x), \ldots, f_l(x))$, $x \in \mathcal{K}$; $\mathcal{K}^{eff}$, $\mathcal{K}^{peff}$ be sets of efficient and properly efficient points of the problem (6). If $\mathcal{K}$ is a convex set, $f_i(x)$, $i = 1, \ldots, l$ are convex functions on $\mathcal{K}$, then

$$f(\mathcal{K}^{peff}) \subset f(\mathcal{K}^{eff}) \subset \bar{f}(\mathcal{K}^{peff})$$

($\bar{f}(\mathcal{K}^{peff})$ denotes the closure set of $f(\mathcal{K}^{peff})$.

2.2 Multi–Objective Stochastic Optimization Problems

Setting

$$f_i(x) = \mathsf{E}_F g_i(x, \xi),\ i = 1, \ldots, l, \quad \mathcal{K} = X_F, \tag{8}$$

then evidently, under assumptions of convex functions $g_i(x, \xi)$, $i = 1, \ldots, l$ on convex, nonempty set X_F, we can (employing Proposition 1) to obtain the set of properly efficient points of the problem (5). According to Remark 2 this set approximate the set of efficient points of (5). Consequently it is suitable, first, to suppose X be nonempty convex set and to analyze properties of the sets type X_F separately:

1. $X_F = X$. In this case $X_F = X$ is a convex set.
2. $X_F = \{x \in X : \mathsf{E}_F g_j^*(x, \xi) \leq 0,\ j = 1, \ldots,\ l'\}$. Evidently if $g_j^*(x, \xi),\ j = 1, \ldots, l'$ are convex functions on convex nonempty set X, then X_F is a convex set.
3. X_F given by individual probability constraints. In particular, in this case we assume $l' = s$, there exist functions $\bar{g}_j(x)$, $j = 1, \ldots, s$ defined for $x \in X$ such that $g_j^*(x, z) = \bar{g}_j(x) - z_j$, $j = 1, \ldots, l$ and

$$X_F := X_F(\alpha) := \bigcap_{j=1}^{s} \{x \in X :\ P[\omega : \bar{g}_j(x) \leq \xi_j] \geq \alpha_j\},$$

where $\alpha_j \in (0,\,1)$, $j = 1,\,\ldots,\,s$, $\quad \alpha = (\alpha_1,\,\ldots,\,\alpha_s)$, $\quad z = (z_1,\,\ldots,\,z_s)$.

If $\bar{g}_j(x)$, $j = 1,\,\ldots,\,s$ are convex functions, P_F absolutely continuous with respect to the Lebesque measure on R^s, then X_F is a convex set (for more details see, e.g., [9]).

The situation with general probability constraints is rather complicated; see, e.g., [16].

4. To deal with the last case, let $l' = 1$, $g_1^*(x,\,\xi)$ be a real-valued function, $Y(\xi)$ a random value. If for $x \in X$ there exists finite $\mathsf{E}_F g_1^*(x,\xi)$, $\mathsf{E}_F Y(\xi)$ and if

$$F^2_{g_1^*(x,\,\xi)}(u) = \int_{-\infty}^{u} F_{g_1^*(x,\,\xi)}(y)dy, \quad F^2_Y(u) = \int_{-\infty}^{u} F_Y(y)dy, \quad u \in R^1,$$

then we can define the second order stochastic dominance constraints X_F by

$$X_F \,=\, \{x \in X : F^2_{g_1^*(x,\,\xi)}(u) \le F^2_Y(u) \quad \text{for every} \quad u \in R^1\}. \tag{9}$$

Employing the results [14], the stochastic second order dominance constraints (9) can be rewritten in a more friendly form:

$$X_F = \{x \in X : \mathsf{E}_F(u - g_1^*(x,\xi))^+ \le \mathsf{E}_F(u - Y(\xi))^+ \quad \text{for every} \quad u \in R^1\}. \tag{10}$$

If $g_1^*(x,\,\xi)$ is a concave function on X, then X_F is a convex set.

Since for the optimization problem given by (8), (10) the Slater's condition is not fulfilled generally, it is necessary to relax constraints set (for more details see, e.g., [2], [11]).

The multi–objective problem given by (8) (with X_F fulfilling first and third case, $g_i(x,\,\xi)$, $i = 1,\,\ldots,\,l$ be strongly convex function) has been investigated in [10], where also the definition of a strongly convex function can be found. However, [10] is mainly focus on the case when the theoretical measure P_F is replaced by empirical one given by independent random sample. To obtained these results the stability with respect to the Wasserstein metric has been there investigated.

The special case of the functions f_i has been considered in [5]. In particular, there were considered the following multi–objective two-stage stochastic problem:

$$\begin{aligned} &\text{Find} \quad \min f_i(x) \,=\, g_i'(x) + \mathsf{E}_F \min q'y,\; i = 1,\,\ldots,\,l \\ &\text{s.t.} \quad Ax = b, \\ &\qquad\; Dx + Wy = \xi, \quad x \ge 0, \quad y \ge 0. \end{aligned}$$

A, D, W, b, q are deterministic matrix of the corresponding dimensions, g_i', $i = 1,\,\ldots,\,l$ suppose to be linear deterministic.

The stability of the last problem considered with respect P_F and based on the bounded Lipschitz metric, has been investigated in [5]. We recall this work because it has been first one dealing with the stability of multi–objective stochastic problem.

Remark 4. An idea of scalarization has been also employed in [8]. However there the approach is combined with utility function approach. Consequently, there the linear dependence objectives on the probability measure is a very suitable property. In [10] this property can be replaced by more general assumptions.

3 Multi–Objective Stochastic Objectives via Stochastic Dominance

The relations (9), (10) recalled the constraints set given by the second order stochastic dominance. Considering in (1) the case $l = 1$ and g_1^*, $j = 1,\,\ldots,\,l'$ being omitted, we can evaluate the solution x by the stochastic dominance. To this end, first, we generalize the problem (3) and consider the problem:

$$\text{Find} \quad \max \mathsf{E}_F g_1(x,\,\xi), \quad \min \rho(g_1(x,\,\xi))) \quad \text{s.t.} \quad x \in X, \quad \rho(\cdot) \quad \text{denotes a risk measure.} \tag{11}$$

(11) is two objectives optimization problem. Evidently to optimize simultaneously both objectives is very often impossible (see also a comment to the problem (3)). Following the approach of Markowitz we can obtain the problem:

$$\text{Find} \quad \max\{(1-\lambda)\mathsf{E}_F[g_1(x,\,\xi)] - \lambda\rho(g_1(x,\,\xi)\} \quad \text{s.t.} \quad x \in X; \quad \lambda \in \langle 0,\,1\rangle. \tag{12}$$

To recall examples of the risk measure ρ considered in [8] we set $V := g_1(x,\,\xi)$. They are

1. variance – $\rho(V)(:= \mathrm{var}(V)) = \mathsf{E}_F[V - \mathsf{E}_F V]^2$,
2. absolute semi–deviation – $\rho(V)\,(:= \bar{\delta}(V)) \;=\; \mathsf{E}_F[\max\,(\mathsf{E}_\mathsf{F}[V] - V),\,0]$,
3. the standard semi–deviation – $\rho(V)\,(:=\; (\delta(V)) \;=\; (\mathsf{E}_F[(\max(\mathsf{E}_F[V] - V,\,0))^2])^{1/2}$,
4. $\rho(V)\,(:= \text{Average Value} - \text{at} - \text{Risk}) = AV@R_\alpha(V) \quad$ for some fixed $\quad \alpha \in [0,\,1]$.
   (For the definition of $AV@R_\alpha(V)$ see, e.g., [8]), [19].)

It is well known that the second order stochastic dominance corresponds to order in the space of non-negative nondecreasing concave utility functions. If we denote by the symbol $\succeq_2$ second order stochastic dominance, then the following definition can be found in [8], [15].

Definition 3. The mean–risk model (11) is called consistent with the second order ($\succeq_2$) stochastic dominance if for every $x \in X$ and $y \in X$,

$$g_1(x,\,\xi) \;\succeq_2\; g_1(y,\,\xi) \quad \Longrightarrow \quad \mathsf{E}g_1(x,\,\xi) \;\geq\; \mathsf{E}g_1(y,\,\xi) \quad \text{and} \quad \rho(g_1(x,\,\xi)) \;\leq\; \rho(g_1(y,\,\xi)).$$

It has been recalled in [8] that the mean–risk model using Average Value–at–Risk at some level α is consistent with second order stochastic dominance relation.

Remark 5. The Markowitz mean-variance model is not consistent with second order stochastic dominance relation ($\succeq_2$), so it is not perfectly suited as a decision aid for rational, risk averse decision marker. In the case of absolute semi–deviation and standard semi–deviation the situation is a little bit more complicated.

Till now we have employed univariate second order stochastic dominance order. However, there are known also (from the literature) the definitions of multivariate stochastic orders. Evidently they are defined on sets of vector random variables. There are known approaches those generalize univariate stochastic order to multivariate case. To recall them let $\mathbf{X}$, $\mathbf{Y}$ be two m–dimensional vectors with components $(X_1,\,\ldots,\,X_m), \quad (Y_1,\,\ldots,\,Y_m)$.

Definition 4. $\mathbf{X}$ dominates $\mathbf{Y}$ in sense of components in second order if $X_j \;\succeq_2\; Y_j \quad$ for every $\quad j = 1,\,\ldots,\,m$.

Definition 5. $\mathbf{X}$ dominates $\mathbf{Y}$ in sense of positive linear second order if $\mathbf{a^T X} \;\succeq_2\; \mathbf{a^T Y}$ for all $\mathbf{a} \geq 0$.

The others generalized models of the multivariate stochastic dominance have appeared in the last time. We recall a work [17] devoted very carefully to this topic.

Applications of multivariate stochastic orders to multi–objective stochastic programming problems can be found in [8].

4 Special Approach

At the end we mention one special approach in which an objective and constraints are determined by one decision parameter (see [12]). The problem can be introduced in the form:

$$\begin{aligned} &\text{Find} \quad \max q^T x + \eta(p) \\ &\text{s.t.} \quad\; Ax \geq b, \\ &\qquad\; P\{Tx \leq d\} \geq p, \quad p \geq \underline{p}, \quad 0 \leq x \leq u. \end{aligned} \tag{13}$$

$T(:= T(t_{i,j}))$ is a matrix with rows $T_1^T, \dots . T_s^T$ discretely distributed random vectors not necessary independent), moreover, each component of $T_i = t_{i,j}\xi_j$, where $t_{i,j}$ is a scalar and ξ_j random variable; q, A, b, u, d, $\underline{p}$ are given deterministic with suitable dimension; $\eta(p)$ is monotone increasing function of p. Evidently the problem (13) can find applications in production planing; $q^T x$ can be considered as a profit of production, $\underline{p}$ can be interpreted as the lowest acceptable reliability level of quality control process or the ready rate service level provided to customers.

5 Conclusion

We have tried to give a brief survey of the approaches to multi–objective problems with a random element. These problems arise in applications, some of them can be found in [3], [10], [18]. However, to deal with them is beyond of the scop of this contribution .

Acknowledgment This work was supported by the Czech Science Foundation under grant 18-02739S.

References

[1] Birge, J.R. and Louveaux, F.: *Introduction in Stochastic Programming.* Springer, Berlin 1999.

[2] Dentcheva, D. and Ruszczynski, A.: Optimization with stochastic dominance constraints. *SIAM Journal of Optimization* **14** (2003), 548–566.

[3] Dupačová, J., Hurt, J. and Štěpán, J. *Stochastic Modeling in Economics and Finance.* Kluwer,Dortrecht 2002.

[4] Caballero, R., Cerdá, F., Munoz, M. M., Rey, L., and Stancu–Minasian, I. M.: Efficient solution concepts and their relations in stochastic multiobjective programming. *Journal of Optimization Theory and Applications.* **110** (2015), 1, 53–74.

[5] Cho, G-M.: Stability of the multi objective linear stochastic programming problems. *Bull. Korean Math. Soc.* **32** (1995), 2, 287–296.

[6] Ehrgott, M.,: *Multicriteria Optimization.* Springer, Berlin 2005.

[7] Geoffrion, A. M.: Proper efficiency and the theory of vector maximization. *J. Math. Anal. Appl.* **22** (1968), 3, 618–630.

[8] Gutjahr, W.J., and Pichler, A.: Stochastic multi–objective optimization: a survey on non–scalarizing methods. *Annals of Operations Research* (2013).

[9] Kaňková, V. On the Stability in Stochastic Programming: the Case of Individual Probability Constraints. *Kybernetika* **33**, 5, 525–544.

[10] Kaňková, V.: A Remark on Multi–Objective Stochastic Optimization with Strongly Convex Functions.*CEJOR* **24**, 309–333.

[11] Kaňková, V., and Omelčenko, V.: Stochastic optimization problem with second order stochastic dominance constrains via Wasserstein metric. *Kybernetika* (submitted).

[12] Lejeune, M. A. and Shen, S.: Multi–Objective Constrained Programming with Variable Risk: New Methods and Applications. *SPEPS* (2014).

[13] Markowitz, H. M.: Portfolio selection. *J. of Finance* **6** (1952), 77–91.

[14] Ogryczak, W. and Rusczynski, A.: From the stochastic dominance to mean–risk models: semidevations as risk measure. *European J. oper. Res.* **116** (1999), (33–50.

[15] Ogryczak, W. and Rusczynski, A.: On consistency of stochastic dominance and mean–semideviation models. *Math. Program. ser.B.* **89** (2001), 217–232.

[16] Prékopa, A.: *Stochastic Programming.* Akadémia Kiadó, Budapest and Kluwer, Dordrecht 1995.

[17] Petrová, B.: Multistage portfolio optimization with multi–variate dominance constraints. *Computational Management Science* (submitted).

[18] Stancu–Minasian, I.M.: *Stochastic Programming with Multiple Objective Functions.* D. Reidel Publishing Company, Dordrecht 1984.

[19] 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. DOI:10.1137/1.9780898718751.

How to down-weight observations in robust regression: A metalearning study

Jan Kalina[1], Zbyněk Pitra[2]

Abstract. Metalearning is becoming an increasingly important methodology for extracting knowledge from a data base of available training data sets to a new (independent) data set. The concept of metalearning is becoming popular in statistical learning and there is an increasing number of metalearning applications also in the analysis of economic data sets. Still, not much attention has been paid to its limitations and disadvantages. For this purpose, we use various linear regression estimators (including highly robust ones) over a set of 30 data sets with economic background and perform a metalearning study over them as well as over the same data sets after an artificial contamination.
We focus on comparing the prediction performance of the least weighted squares estimator with various weighting schemes. A broader spectrum of classification methods is applied and a support vector machine turns out to yield the best results. While results of a leave-1-out cross validation are very different from results of autovalidation, we realize that metalearning is highly unstable and its results should be interpreted with care. We also focus on discussing all possible limitations of the metalearning methodology in general.

Keywords: metalearning, robust statistics, linear regression, outliers.

JEL classification: C14
AMS classification: 68T37

1 Introduction

Metalearning can be characterized as a perspective methodology for extracting knowledge from a data base of training data sets, allowing to apply the knowledge to new independent (validation) data sets. It can be described as learning to learn over metaknowledge, i.e. knowledge about whole data sets which serve as a prior knowledge rather than measured values contained in these data sets. Metalearning represents an approach to machine learning (i.e. automated statistical learning) popular in recent computer science and data mining [2], starting to penetrate also to economic applications [1] suitable also (but not limited to) high-dimensional economic data [14].

The currently most renowned works on metalearning principles [2, 13] recommend metalearning especially for those domains, in which a theoretical knowledge would be hard to acquire. However, besides appealing properties of metalearning, we must also realize its limitations, mainly its instability and sensitivity to data contamination. However, a truly critical evaluation of metalearning seems to be still missing. It is mainly the fully automatic character of the metalearning process which hinders a profound interpretation of the results. The community of computer scientists finds however heuristics and black-box procedures more appealing, i.e. does not incline to detailed interpretations anyway.

Metalearning principles have been already applied to search for the best robust regression method in [11]. Our new approach here is to use metalearning to predict the most suitable weighting scheme for the highly robust least weighted squares estimator of [15]. In addition, we accompany the study with a larger comparison of different classification methods from both multivariate statistics and machine learning. First, we recall principles and methods of robust regression in Section 2. We also propose several new weighting schemes for the least weighted squares estimator. Our metalearning study is described in Section 3 and the results are presented in Section 4. A discussion follows in Section 5. The final Section 6 discusses the instability of metalearning in general.

2 Robust regression

Throughout this paper, the standard linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + e_i, \ldots, i = 1, \ldots, n, \tag{1}$$

[1]Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, Praha 8 & Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 2, Praha 8, Czech Republic, kalina@utia.cas.cz
[2]Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 2, Praha 8, Czech Republic, pitra@cs.cas.cz

is considered with n values of p regressors and a continuous response Y under the presence of random errors $e_1, \dots, e_n$. The least squares estimator, which is the notoriously known standard estimation tool for the linear regression model, is however too vulnerable to the presence of outlying measurements (outliers) [7].

Within the framework of robust statistic, a variety of robust estimators was proposed as an alternative to the least squares. These robust tools are more suitable estimation techniques for data contaminated by outliers and thus gradually become important for the analysis of economic data [6]. M-estimators are the most commonly used robust methods, but do not possess a high breakdown point which has become of one fundamental robustness measures [4]. Therefore, highly robust estimators were proposed as an alternative, which is more reliable under a more severe data contamination.

The least weighted squares (LWS) estimator of [15] represents a (possibly highly) robust estimator defined as

$$\arg\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} w_i u^2_{(i)}(b), \tag{2}$$

where $u_i(b)$ is a residual corresponding to the i-th observation for a given b,

$$u^2_{(1)}(b) \leq u^2_{(1)}(b) \leq \dots \leq u^2_{(n)}(b). \tag{3}$$

are values arranged in ascending order and h is a given trimming constant. Its properties were investigated by [3] or [9]. The estimator is consistent and asymptotically normal for $n \to \infty$. If suitable weights are chosen, the estimator is efficient under normal (non-contaminated) models and at the same time robust under contamination. This property seems to perform well on real data sets in various applications [8, 10].

Nevertheless, a systematic comparisons of the performance of robust regression estimator is missing in spite of intensive attempts [12], while proving theoretical results is tedious and only of limited application, because the results often depend on unknown quantities. Thus, there remains a lack of comparisons of robust estimators on real data and particularly it remains unknown how to choose weights for the LWS for a given data set. Indeed, insufficient comparisons can be described as one of main reasons why robust methods have not much penetrated to real applications. Therefore, metalearning seems as a reasonable tool for the task of predicting the most suitable method for a particular (new) data set or under particular conditions.

3 Description of the study

We proposed and performed a metalearning study with the aim to compare various linear regression estimators and to find a classification rule allowing to predict the best one for a given (new) data set. Based on a data base of training data sets, the aim is also to detect the most relevant criteria for determining the most suitable weights.

The primary learning task is to fit various linear regression estimators for each of the given data sets. The best estimator is found using a specified characteristic of a goodness of fit. The subsequent metalearning part has the aim to learn a classification rule allowing to predict the best regression method for a new data set not present in the training data base. Its input data are only selected features of individual data sets together with the result of the primary learning, which typically has the form of the index of the best method for each of the training data sets.

In general, the user of metalearning must specify a list of essential components (parameters) [13], which will be now described together with our choices for the particular metalearning study of Section 4.

3.1 Primary learning

Metalearning should always use real data sets, because any random generation of data is performed in a too specific (i.e. non-representative, biased) way. We use 30 data sets, while 24 of them were used already by [11] and we include 6 additional data sets with a clearly economic motivation, which come from the online UCI Repository. The list of all these 30 publicly available data sets is presented in Table 2.

In each of the data sets, we consider the model (1) and use one of the following five estimators, where (4) and (5) are novel proposals of this paper:

1. Least squares.
2. LWS with data-dependent adaptive weights of [3].
3. LWS with linear weights
$$w_i = \frac{2(n+1-i)}{n(n+1)}, \quad i = 1, \dots, n, \tag{4}$$
4. LWS with trimmed linear weights. Let the true level of contamination equal $\varepsilon \cdot 100\ \%$ for a known $\varepsilon \in [0, 1/2)$.

In this paper, we take $h = \lfloor 3n/4 \rfloor$, where $\lceil x \rceil = \min\{n \in \mathbb{N};\ n \geq x\}$. The weights equal

$$w_i = \frac{h - i + 1}{h} \mathbb{1}[i \leq h], \quad i = 1, \ldots, n, \tag{5}$$

where $\mathbb{1}[.]$ denotes an indicator function.

5. LWS with weights generated by the (strictly decreasing) logistic function

$$w_i = \left(1 + \exp\left\{\frac{i - n - 1}{n}\right\}\right)^{-1}, \quad i = 1, \ldots, n. \tag{6}$$

In addition, each choise of weights is standardized to fulfil $\sum_{i=1}^{n} w_i = 1$. For the prediction measure, we use the most standard choice, i.e. the (prediction) mean square error (MSE). In the primary learning task, we find the best method for each data set. This is done using MSE in a leave-1-out cross validation, which represents a standard attempt for an independent validation. There is 1 randomly chosen observation left out, the estimator is computed and used to predict the value for the observation being left out. This is repeated for all n observations and the resulting values of the MSE are averaged. Then, the output of the primary learning is the knowledge (i.e. factor variable, index) of the best method for each of the data sets.

3.2 Metalearning

The subsequent metalearning task exploits 10 features for each data set and the factor variable of Table 1 denoting the index of the best method. We use the following features.

1. The number of observations n,
2. The number of regressors p (excluding the intercept),
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 (by the subjective method of Section 4 of [9]),
9. Heteroscedasticity of residuals evaluated as the p-value of the Whites test,
10. Condition number of the matrix $(X^T X)^{-1}$.

For the subsequent metalearning task, which is a task of classification to 5 groups, we exploit various classification methods. For those implemented in R software, we use default settings of all their parameters. This is true for support vector machines (SVM), k-nearest neighbors, a classification tree, and others. We use several less known methods including a regularized version of linear discriminant analysis (LDA) denoted as SCRDA of [5] or a robust version of LDA denoted as quadratic MWCD classification [8]. We also investigate the effect of dimensionality reduction, particularly in the form of replacing the data by 3 principal components obtained by the principal component analysis (PCA).

4 Results

We used the R software for all the computations. Table 1 shows the best method for each of the data sets using the notation $1, \ldots, 5$ according to list of methods in Section 3.1. Using MSE as a measure of prediction performance, both autovalidation and leave-1-out cross validation were used, where autovalidation means predicting for observations present in the training data, while cross validation is a standard attempt for an independent validation. It seems from the autovalidation study that the least squares estimator is the most successful among the five possibilities, while the LWS with data-dependent weights of [3] is the most successful according to the cross validation.

Further, the training 30 data sets are classified to one of the 5 groups using the 10 features. The results for various classifiers are overviewed in Table 2, namely as classification accuracy of the metalearning classification task evaluated in a leave-1-out cross validation study. The classification accuracy is defined as the number of correctly classified cases divided by the total number of cases (i.e. percentage of correct results). If all 10 features are used, the best result is 0.40 obtained with an SVM classifier with a Gaussian kernel.

5 Discussion

Let us now discuss the performance of robust methods withing the metalearning study aiming to predict the best weighting scheme for the highly robust LWS estimator in linear regression.

The LWS estimator turns out to be quite often more suitable than least squares, which is a novel argument in favor of the method. It is an important task to decide, for which there have been no recommendations for a practical data analysis [15]. The data-depent weights of [3] turn out to be the winner of the comparisons. However, it is not the best option for all data sets and the metalearning study allows to predict the most suitable weighting scheme also for a new (independent) data set.

It cannot be easily interpreted which features are the most relevant for predicting the most suitable weighting scheme. The reason is the black-box character of the SVM, which turns out to be best classifier here. Some other classifiers, which are quite commonly used in metalearning, yield relatively weaker results. This is also the case of the k-nearest neighbors, which seem to perhaps the most common method in the metalearning task. Further, our study presents also a unique comparison of SCRDA and MWCD-LDA with standard LDA. Both these approaches are able to outperform LDA, which is especially for the relatively recent MWCD-LDA a new argument in its favor.

The effect of reducing the dimensionality, which is often performed with metalearning, is clearly leading to a too drastic loss of information. Here, the principal component analysis does not allow to construct a reliable subsequent classification rule.

Finally, a big different between results of autovalidation and cross validation reveal the instability of metalearning, because omitting a single data set from the training data base leads to huge differences in the performance of the resulting classification rule. Therefore, we include a final Section 6 discussing the instability of metalearning in general and its reasons.

6 Instability of metalearning

Besides presenting results of a particular metalearning study across 30 real publicly available data sets, our study reveals also limitations of metalearning and motivates a possible future critical evaluation of the metalearning process. Possible factors contributing to the sensitivity and/or instability of metalearning include:

- The choice of data sets. We use here a rather wide spectrum of data sets with different characteristics from different research tasks, while metalearning is perhaps more suitable only for more homogeneous data (e.g. with analogous dimensionality) or for data from a specific narrow domain. Apparently, it remains difficult (and unreliable) to perform any extrapolation for a very different (outlying) data set.
- The prediction measure. In our case, MSE is very vulnerable to outliers.
- The number of methods. If their number is larger than very small, we have the experience that learning the classification rule becomes much more complicated and less reliable.
- The classification methods for the metalearning task depend on their own parameters or selected approach, which is another source of uncertainty and thus instability.
- Solving the metalearning method by classification tools increases the vulnerability as well, because only the best regression estimator is chosen ignoring information about the performance of other estimators.
- The process of metalearning itself is too automatic so the influence of outliers is propagated throughout the process and the user cannot manually perform an outlier detection or deletion.

A less sensitive (more robust) approach to metalearning remains to represent an important topic for future research. It is clear that such alternative methodology requires a more complex effort than just a robustification of each of its individual steps (e.g. using a robust prediction measure, a robust classifier or a robust dimensionality reduction). One idea seems to consider a robust version of the mean square error. Robustifying the final classification may be performed by means of using ensemble classifiers or perhaps regression methodology. This would require replacing the right column from Table 1 by additional knowledge, e.g. the performance of all 5 estimators for each data set. It is also important to investigate the effect of data contamination as well as dimensionality reduction on the sensitivity of metalearning.

Acknowledgements

The research was supported by grants 17-07384S and 17-01251S of the Czech Science Foundation.

References

[1] Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). MetaFraud: A meta-learning framework for detecting financial fraud. *MIS Quarterly*, 36, 1293–1327.

[2] Brazdil, P., Giraud-Carrier, C., Soares, C., & Vilalta, E. (2009). *Metalearning: Applications to data mining.* Berlin: Springer.

[3] Čížek, P. (2011). Semiparametrically weighted robust estimation of regression models. *Computational Statistics & Data Analysis*, 55 (1), 774–788.

[4] Davies, L. & Gather, U. (2005). Breakdown and groups. *Annals of Statistics*, 33, 977–1035.

[5] Guo, Y., Hastie, T., & Tibshirani, R. (2007). Regularized discriminant analysis and its application in microarrays. *Biostatistics*, 8 (1), 86–100.

[6] Hubert, M., Rousseeuw, P.J., & van Aelst, S. (2008). High-breakdown robust multivariate methods. *Statistical Science*, 23, 92–119.

[7] Jurečková, J., Sen, P.K., & Picek, J. (2012). *Methodology in robust and nonparametric statistics.* Boca Raton: CRC Press.

[8] Kalina, J. (2012): Highly robust statistical methods in medical image analysis. *Biocybernetics and Biomedical Engineering*, 32 (2), 3–16.

[9] Kalina, J. (2015). Three contributions to robust regression diagnostics. *Journal of the Applied Mathematics, Statistics and Informatics*, 11 (2), 69–78.

[10] Kalina, J. (2015). A robust supervised variable selection for noisy high-dimensional data. *BioMed Research International*, 2015, Article 320385.

[11] Kalina, J. & Peštová, B. (2017). Robust regression estimators: A comparison of prediction performance. In *Conference Proceedings MME 2017, 35th International Conference Mathematical Methods in Economics* (pp. 307–312). Hradec Králové: University of Hradec Králové.

[12] Mašíček, L. (2004). Optimality of the least weighted squares estimator. *Kybernetika*, 40, 715–734.

[13] Smith-Miles, K., Baatar, D., Wreford, B., & Lewis, R. (2014). Towards objective measures of algorithm performance across instance space. *Computers & Operations Research*, 45, 12–24.

[14] Varian, H.R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28, 3–28.

[15] Víšek, J.Á. (2011). Consistency of the least weighted squares under heteroscedasticity. *Kybernetika*, 47, 179–206.

[16] Yohai, V.J. & Zamar, R.H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, 83, 406–413.

Classification method	Classification accuracy
Classification tree	0.27
Logistic regression	0.33
LDA	0.33
SCRDA	0.37
Quadratic MWCD classification	0.37
Multilayer perceptron	0.30
k-NN ($k = 3$)	0.27
SVM (linear)	0.37
SVM (Gaussian kernel)	0.40
PCA $\Longrightarrow$ LDA	0.23
PCA $\Longrightarrow$ Logistic regression	0.23

Table 1 Results of metalearning evaluated as the classification accuracy in a leave-1-out cross validation study.

Index	Data set	The best method found by autovalidation	LOOCV
1	Aircraft	1	2
2	Amazon access samples	2	5
3	Ammonia	1	2
4	Auto MPG	1	2
5	Cirrhosis	3	3
6	Coleman	1	1
7	Communities and crime	2	5
8	Delivery	1	2
9	Education	1	2
10	Electricity	1	3
11	Employment	1	2
12	Energy efficiency	2	5
13	Facebook metrics	2	5
14	Furniture 1	1	1
15	Furniture 2	1	3
16	GDP growth	5	2
17	Houseprices	2	2
18	Housing	1	3
19	Imports	1	4
20	Insurance company benchmark	4	4
21	Istanbul stock exchange	2	5
22	Kootenay	3	3
23	Livestock	1	3
24	Machine	1	2
25	Murders	1	2
26	NOx	1	2
27	Octane	1	1
28	Pasture	1	4
29	Pension	2	3
30	Petrol	1	2

Table 2 Results of primary learning in a leave-1-out cross-validation study (LOOCV).

Cost Characteristics of EWMA Based AOQL Variables Sampling Plans

Nikola Kaspříková[1]

Abstract. The acceptance sampling plans minimizing the mean inspection cost per lot of the process average quality when the remainder of the rejected lots is inspected were designed by Dodge and Romig for the inspection by attributes. Plans for the inspection by variables were then proposed, including a procedure based on the EWMA statistic. The plans for the inspection by variables may be more economical than the corresponding attributes sampling plans. The recently proposed plans using EWMA statistic in the decision procedure may lead to further improvements in the inspection cost. The design of the EWMA-based rectifying AOQL sampling plans minimizing the mean inspection cost per lot of process average quality is recalled and the evaluation of the comparative economic efficiency of this EWMA-based plan is shown, using a simple economic model. For the comparison of the plans and for making the decision on which plan to use in a particular situation in practice, we propose to consider a cost characteristic which may be calculated for the plan easily.
The EWMA-based rectifying AOQL sampling plans are calculated using an R software extension package.

Keywords: acceptance sampling, cost, AOQL, EWMA.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Sampling inspection plans are basic industry statistics tools for quality control and there exist many types of sampling inspection plans. Dodge and Romig (see e.g. [2]) have designed the average outgoing quality limit (AOQL) sampling plans minimizing the mean inspection cost per lot of process average quality when the remainder of rejected lots is inspected. The plans were originally designed by Dodge and Romig 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 [6]. 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 exponentially weighted moving average (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 in comparison with the plans without memory, as it allows using information on the quality in the previous lots. The EWMA-based Lot Tolerance Percent Defective (LTPD) plans for the unknown standard deviation case are introduced in [7]. 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 [5] and a simple economic model has been used for the evaluation of economic efficiency of the plans. In this paper we introduce new cost parameter, the break-even value of cost ratio, for the EWMA-based AOQL plans. The design of this value follows the idea used in [8] for the AOQL plans plans without memory. This value is an interesting characteristic of the particular situation in practice and may be used to guide the business decisions regarding which plan to use (i. e. the plan for the inspection by attributes or the EWMA-based plan for the inspection by variables and attributes).

The structure of this paper is as follows: first, the design of the original AOQL sampling plans for the inspection by attributes, as introduced by Dodge and Romig (see [2]), is recalled. Then we recall the design of the the AOQL variables sampling plans based on the usage of the EWMA statistic in the inspection procedure and also recall the simple economic model used to asses the comparative economic efficiency of the plans. Then we introduce the new

[1]University of Economics in Prague, Department of Mathematics,
Nám. W. Churchilla 4, Praha, Czech Republic, nb33@tulipany.cz

cost parameter and discuss its usage. For an example situation we calculate the plan in the R software extension package [6] which implements the algorithms for the calculation of the plans and which has been published on the Comprehensive R Archive Network and search for the value of the new cost parameter.

2 Dodge-Romig 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.

3 Sampling plans for the inspection by variables

As an alternative to the sampling plans for the inspection by attributes, the AOQL plans for the inspection by variables and attributes based on the EWMA statistic for the case of the unknown standard deviation are designed. The operating characteristic used in [7] for the design of the LTPD plans is applied here. The 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 efficiency, 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 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 should be estimated based on the real cost calculation in practice. Usually it is $c_m > 1$.

3.1 Economic evaluation of the plans

For the comparison of the plans - the EWMA plan for inspection by variables and attributes (n, k) and the Dodge-Romig plan for the inspection by attributes (n_a, c) from an economic point of view we can use a simple economic model (see [5]) 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:

$$c = I_{ms}(n, k) - I_s(n_a, c) + f, \tag{12}$$

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 usually 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.

We denote the break-even value of c_m parameter as cm_b and define this parameter as follows: cm_b is the value of c_m for which the cost difference c equals zero. This value can be calculated easily for a particular situation in practice using numerical methods. If the real value of the quotient of the cost of inspecting an item by variables and the cost of inspecting the item by attributes c_m is lower than cm_b, then one should use the sampling plans for the inspection by variables and attributes. For the situations when c_m is greater than cm_b, the original Dodge-Romig plans are more economical.

4 Example of calculation and economic evaluation of the plans

4.1 Calculation of the plan with LTPDvar package

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 [6] 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).

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 2.5 times higher than the cost of inspecting the item by attributes, so the parameter c_m equals 2.5. 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 [6] 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=2.5,method="ewma2", lam=0.9)
```

and the output of the calculation is as follows.

```
An object of class "ACSPlan"
Slot "n":
[1] 59

Slot "k":
[1] 1.914631
```

So the resulting solution is $n = 59$, $k = 1.914631$.

For the values of the input parameters given in our problem, there is plan (165, 4) for the acceptance sampling by attributes in [2].

4.2 Economic evaluation of the plan

The resulting sampling plan from Example 1 will be evaluated with regard to the economic characteristics and compared with the corresponding AOQL attributes sampling plan as discussed in [2].

For the comparison of these two plans from Example 1 - the EWMA plan for inspection by variables and attributes ($n = 59$, $k = 1.914631$) and the Dodge-Romig plan ($n = 165$, $c = 4$) from an economic point of view we have:

$$c = I_{ms}(59, 1.914631) - I_s(165, 4) + f. \tag{13}$$

The I_{ms} function is implemented in the LTPDvar package too and (for plan $n = 59$, $k = 1.914631$) may be called as shown below.

```
Ims(n= 59, k=1.914631, N= 3500, pbar=0.01, cm=2.5, type="ewma2", lam=0.9)
```

In this case we obtain $I_{ms} = 231$.

Let's suppose now that in our example situation it is $f = 5\,c_s^*$, 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 $c = -4$, 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 c_m parameter 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 c_m equals 2.6, then we get sampling plan for the inspection by variables and attributes as follows:

```
planAOQL(N=3500, pbar=0.01, pL=0.015, cm=2.6,method="ewma2", lam=0.9)
```

and the output of the calculation is as follows.

```
An object of class "ACSPlan"
Slot "n":
[1] 58

Slot "k":
[1] 1.913999
```

So the resulting solution is $n = 58$, $k = 1.913999$.

We get the value of the I_{ms} function for the plan using the following call.

```
Ims(n= 58, k=1.913999, N= 3500, pbar=0.01, cm=2.6, type="ewma2", lam=0.9)
```

In this case we obtain $I_{ms} = 237$.

Then the value of the cost measure c is now 1. 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.

Higher values of the c_m parameter support using rather the attribute sampling plans, as the cost of inspecting the item by attributes is considerably lower than the cost of inspecting the item by variables. For the values of the input parameters in our example, the cm_b value, i. e. the break-even value of the cost parameter c_m, equals 2.572. For the original value of the c_m parameter ($c_m = 2.5$), it holds

$$c_m < cm_b,$$

which means, the plans for the inspection by variables and attributes are more economical than the original plans for the inspection by attributes.

If the cm_b value is near one, then one could make the decision on whether to use the Dodge-Romig plan or the plan for the inspection for variables and attributes even without precisely specifying the value of c_m. If we get the cm_b value near one and have some idea that the real c_m value is somewhat higher, then we opt for the attributes sampling plan. This way of making the decision (i. e. without precisely specifying the value of c_m) may be preferable, since the real cost calculations for getting the value of c_m may incur some cost too.

5 Conclusion

In many situations in the business practice, 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 bring significant economic advantages over the corresponding attribute inspection plans. This improvement in economic efficiency can be evaluated using a simple economic model considering the difference in variable and fixed cost between the variables sampling plan and the attributes sampling plan. The break even value of the cost parameter c_m can be of interest when making the decision on whether to use the plan for the inspection by attributes or the plan for the inspection by variables and attributes.

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. (2015). A new lot inspection procedure based on exponentially weighted moving average, *International Journal of Systems Science* **46**, 1392–1400.

[2] Dodge, H., F., Romig, H. G. (1998). *Sampling Inspection Tables: Single and Double Sampling*. New York: John Wiley.

[3] Jennett, W. J., and Welch, B. L.(1939). 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**, 80–88.

[4] Johnson, N. L., and Welch, B. L.(1940). Applications of the Non-central t distribution, *Biometrika* **38**, 362–389.

[5] Kasprikova, N. (2017). EWMA Based AOQL Variables Sampling Plans and Cost Models. In *Mathematical Methods in Economics* (pp. 319–323). Hradec Králové: University of Hradec Králové.

[6] Kasprikova, N. (2016). *LTPDvar: LTPD and AOQL plans for acceptance sampling inspection by variables.* R package version 1.2. http://CRAN.R-project.org/package=LTPDvar.

[7] Kasprikova, N. (2017). Rectifying LTPD Plans Minimizing Mean Inspection Cost Per Lot of Process Average Quality Using EWMA Statistics. *Journal of Testing and Evaluation* **4**, DOI: 10.1520/JTE20150259.

[8] Kasprikova, N., and Klufa, J. (2015). AOQL sampling plans for inspection by variables and attributes versus the plans for inspection by attributes, *Quality technology and quantitative management* **6**, 133–142.

[9] Klufa, J. (1997). Dodge-Romig AOQL single sampling plans for inspection by variables, *Statistical Papers* **38**, 111–119.

[10] R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Analysis of the phenomenon of long-term memory in financial time series

Zeug-Żebro Katarzyna[1], Szafraniec Mikołaj[2]

Abstract. The phenomenon of a long-term memory of time series has been the subject of researchers' interest for more than six decades. From the moment when H. E. Hurst described the method of rescaled range (R/S) analysis, it was noticed that the nature of many natural phenomena differs from the random walk model and characterized by the so-called persistence. No wonder that the long-term memory quickly came to the attention of economists and stock market players, who saw in it the ability to "beat the market" and increase their profits. The presence of a long-term memory component in a time series means that even very distant observations exert some influence on subsequent realizations of the process. Usually, this relationship is not particularly strong, but it does exist. Interpreting this phenomenon in the context of financial phenomena, one can come to the conclusion that information that has affected the market many days or even weeks ago may still be important for the current stock quotes.
The main aim of the paper is to conduct a comprehensive analysis for the occurrence and strength of the phenomenon of long-term memory in selected financial time series with used Hurst exponent and local Whittle estimator.

Keywords: phenomenon of long-term memory, rescaled range analysis, Hurst exponent, local Whittle estimator, financial time series.

JEL Classification: C13, G12
AMS Classification: 37M10

1 Introduction

Since the appearance of Peters' work [10], the phenomenon of long-term memory in the context of the shaping of rates of return on global stock exchanges arouses considerable interest of researchers. Also in Polish literature appeared many works related to this topic. It is worth mentioning in this context a study of Czekaj, Woś and Żarnowski [2], Orzeszko [8] or Zeug-Żebro [14]. However, most of the work was based on method of rescaled range introduced by Hurst. Research with the use of Whittle estimators for the Warsaw Stock Exchange was conducted by Gurgul and Wójtowicz [3, 4], but they concerned only the interrelations between the volatility of rate of returns and volume.

The actual participation of the phenomenon of long-term memory in the process of shaping the series of rates of return on the financial markets would carry with him serious implications. They would concern both the theory of information effectiveness and investment strategies used by large financial institutions and small stock exchange players - it might be necessary to verify their assumptions.

The main aim of the paper is to conduct a comprehensive analysis of the occurrence and strength of the phenomenon of long-term memory in selected financial time series using the Hurst exponent and the local Whittle estimator. The studies used time series created from closing prices of five stock exchange indices and fifteen companies listed on the Warsaw Stock Exchange. The data covered the period from 04.01.2005 to 23.03.2018. The calculations were carried out using a program written by the author in the Delphi programming language, in the RCran statistical software and the Microsoft Excel package.

2 Phenomenon of long-term memory

Long-term memory, also called the property of long-term dependence, consists in the existence of significant autocorrelations between even distant observations creating a time series [5].

Another property of the autocorrelation function, which is characteristic for the time series with long-term memory, is the rate of its descent. In the case of short-term dependencies described, for example, a stationary

[1] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, 1 Maja 50, 40-287 Katowice, Poland, katarzyna.zeug-zebro@ue.katowice.pl.
[2] University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland, mikolaj.szafraniec@edu.uekat.pl.

model of AR(1), the autocorrelation function approaches zero at an exponential rate. For time series with a feature of long-term memory, this drop also occurs, however, it is much slower and it is hyperbolic [5]. Kwiatkowski notices that the decisive criterion is the rate of descent of the ACF function and not its value [7]. The phenomenon of long-term memory may also occur in the case of series, for which next estimated values of ACF turn out to be statistically insignificant.

The phenomenon of long-term memory has also had more formal definitions (from a mathematical point of view). One of them we find, among others at Palm's work [9]. Let: $\gamma(h) = \langle y_t, y_{t+h} \rangle$ will be a function of autocovariance for subsequent delays of the *h* stationary process. This process is characterized by the occurrence of a long-term memory phenomenon, if this function fulfills the condition:

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty. \tag{1}$$

In the case of short-term memory processes, such as AR (1) with the parameter $\alpha \in (-1; 1)$, the sum of a similar series is a finite value [7]. In literature, we also find other definitions of long-term memory [3]. The stochastic process has a long-term memory if its spectral density function $f(\lambda)$ fulfills the condition:

$$lim_{\lambda \to 0^+} \frac{f(\lambda)}{c\lambda^{-2d}} = 1, \tag{2}$$

where *c* is a constant, while the d parameter is characterized the process memory. This condition means that for low frequencies close to zero the spectral density of the process goes to infinity. If this condition is met, the autocorrelation function of the process disappears at a hyperbolic rate [3]:

$$lim_{h \to \infty} \frac{\rho(h)}{c_\rho h^{2d-1}} = 1, \tag{3}$$

where *c* is a constant. The parameter *d* defines the character of the process's memory. For $d > 0$, the process has a long memory and is called persistent. It shows a tendency to continue trends and to move away from the average value more than a random walk. For parameter $d = 0$, the series is characterized by a lack of memory; an example of a similar process is white noise. For $d < 0$, the spectral density function of the process goes to zero for low frequencies and increases for high frequencies.. Such a process is characterized by negative autocorrelation of next implementations, greater variability than random walk and has a tendency to return to the average value. Such process called anti-persistent [3]. At this point it is worth mentioning that Hursta h^* exponent is another, often encountered parameter describing the memory of the process. The relationship between *d* and h^* is determined by the following equation:

$$d = h^* - 0.5. \tag{4}$$

The property of long memory is also considered in the context of the dynamics of financial time series. Extensive research in this area was carried out by Peters [10], who used for this purpose the analysis of the rescaled ranger R\S proposed by Hurst. The results indicated that the majority of the series analyzed by Peters were characterized by a significant degree of persistence and an h^* exponent greater than 0.5. Thus, he questioned the hypothesis of effective markets, proposing in exchange a hypothesis about the fractal nature of financial markets. However, it is worth considering whether the R\S analysis used by Peters is an appropriate estimator of the long-term memory parameter. Considerations in this topic will be taken later in this paper.

2.1 Hurst exponent

As mentioned above the Hurst exponent [6] is one of measure that allows for search of phenomenon of the long-term memory. 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 anti-persistent 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 $\{x_1, x_2, ..., x_N\}$ it runs through the following steps [1, 14]:

(1) transform the above time series into $m = N - 1$ logarithmic rates of return:

$$y_k = \log(x_{k+1}/x_k), \quad k = 1, 2, \ldots, N-1. \tag{5}$$

(2) Next, share a series (5) 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:

$$z_{ij} = y_{ij} - \bar{y}_j, \tag{6}$$

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:

$$q_{ij} = \sum_{l=1}^{i} z_{lj}, \qquad i = 1, 2, \ldots, t, \quad j = 1, 2, \ldots, T. \tag{7}$$

(5) The range of the i-th interval is defined as:

$$R_j = \max(q_{ij}) - \min(q_{ij}). \tag{8}$$

(6) Calculate the rescaled range series (R/S):

$$\alpha_{jt} = R_j / S_j, \tag{9}$$

where: $S_j = \sqrt{\frac{1}{t}\sum_{i=1}^{t} z_{ij}^2}$.

(7) Next, calculate:

$$(R/S)_t = (1/T)\sum_{j=1}^{T} \alpha_{jt}. \tag{10}$$

(8) The above procedure is carried out for different lengths of time series t.

(9) To determine the Hurst exponent, we should calculate the logarithm on both sides in the following equation:

$$(R/S)_t = ct^H, \tag{11}$$

where: H is the Hurst exponent, c is a constant, and t is the expected value of the rescaled range:

$$\ln((R/S)_t) = \ln c + H \ln t. \tag{12}$$

(10) 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.

2.2 Local Whittle Estimator

One of estimators used in this work is so called local Whittle estimator, described in Robinson's paper [11]. It is based on the asymptotic properties of the spectral density function $f(\lambda)$ of stationary processes:

$$f(\lambda) \sim G\lambda^{1-2H} \; as \; \lambda \to 0^+, \tag{13}$$

where G is constant that satisfies the condition $G \in (0, \infty)$ and H is a long memory parameter, such that $H \in (0, 1)$. For low frequencies and $H \in (0, \frac{1}{2})$ the value of function $f(\lambda)$ tends to zero, whereas while $H \in (\frac{1}{2}, 1)$ the value of spectral density function tends to infinity. If $H = \frac{1}{2}$ function $f(\lambda)$ tends to a finite positive constant at zero frequency.

Local Whittle estimator is based on the periodogram $I(\lambda)$ of the time series $\{y_t\}$. To calculate estimate of H_0 we have to find the minimum of objective function:

$$Q(G, H) = \frac{1}{m}\sum_{j=1}^{m}\left\{log G\lambda_j^{1-2H} + \frac{\lambda_j^{2H-1}}{G} I_j\right\}, \tag{14}$$

where $I_j = I(\lambda_j)$, and m is integer satisfying $m < n$. If we replace G variable with estimator $\hat{G}$:

$$\hat{G}(H) = \frac{1}{m}\sum_{1}^{m} \lambda_j^{2H-1} I_j, \tag{15}$$

then the objective function (2.2) may be simplified to the function of one variable:

$$R(H) = log\, \hat{G}(H) - (2H-1)\frac{1}{m}\sum_{1}^{m} log\lambda_j \, . \tag{16}$$

Local Whittle estimator belongs to class of semiparametric estimators. Its main advantage is that it doesn't require a model specification. If we choose the parameter m appropriately, the estimation result is not biased even in case of presence of short memory components, such like autoregression or moving average phenomenon. What's more, local Whittle estimator allow us to avoid assumption of the Gaussianity of the studied time series, unlike many others estimators. Under conditions described by Robinson [11], such as:

$$\frac{1}{m} + \frac{m}{n} \to 0 \; as \; n \to \infty, \tag{17}$$

local Whittle estimator is consistent and statistic

$$\sqrt{m}(\hat{H} - H_0) \tag{18}$$

is normally distributed $N(0,\frac{1}{4})$. Taqqu i Teverovsky [13] shown, that when the time series are long enough and we choose appropriate value of parameter m, the local Whittle estimator perform not much worse than the original Whittle method with correct model specification.

3 The purpose and conduct of the study

As part of the study, we estimated a long memory parameter H and the Hurst exponent for 20 selected time series of logarithmic rates of returns. Analysis covered stock of 15 companies and 5 stock market indexes listed on Warsaw Stock Exchange. Each of the time series included 3312 observation for the period of time from 04.01.2005 to 23.03.2018. Chosen time window let us consider market condition both in financial crisis 2007 and periods of stable growth.

Although the Hurst exponent of the random series is equal to 0.5, only for the correspondingly large N values H will approach this limit. Therefore, to draw conclusions about the randomness of the examined series, the Hurst exponent should be compared with the expected value of the exponent of a random series of the same length [12]. The Hurst exponent different from the expected one, testifies to the existence of a series of long-term memory.

The empirical research allowed to determine the Hurst exponent and its expected value (Table 1)[3] where

$$E((R/S)_t) = \frac{t}{(t-0{,}28)\sqrt{\frac{\pi}{2}t}} \sum_{i=1}^{t-1} \sqrt{\frac{t-i}{i}}, \tag{19}$$

and $E((R/S)_t)$ is expected value of Hurst exponent and t is the length of substrings analyzed in the method R/S.

Company / Index	Estimated Hurst exponent	Expected Hurst exponent	Number of observations/ number of divisors
SWIG80	**0.6686**	0.5512	3312/22
INVESTORMS	**0.6603**	0.5512	3312/22
MWIG40	**0.6353**	0.5512	3312/22
AMICA	**0.6195**	0.5512	3312/22
FORTE	**0.6113**	0.5512	3312/22
POLNORD	**0.6070**	0.5512	3312/22
GROCLIN	**0.6023**	0.5512	3312/22
KGHM	**0.5969**	0.5512	3312/22
DEBICA	**0.5901**	0.5512	3312/22
WIG	**0.5885**	0.5512	3312/22
CCC	**0.5824**	0.5512	3312/22
INGBSK	**0.5797**	0.5512	3312/22
RELPOL	**0.5695**	0.5512	3312/22
AGORA	**0.5662**	0.5512	3312/22
CDPROJEKT	**0.5563**	0.5512	3312/22
WIG20	**0.5555**	0.5512	3312/22
BZWBK	**0.5471**	0.5512	3312/22
PKNORLEN	**0.5469**	0.5512	3312/22
PKOBP	**0.5197**	0.5512	3312/22
PEKAO	**0.5014**	0.5512	3312/22

Table 1 Hurst exponent and its expected value: estimation results

The results obtained showed that for most financial time series the Hurst exponent clearly differs from the expected value. This means that these financial series are characterized by long-term memory. Time series of

[3] In order to estimate the Hurst exponent and its expected value, the author's program written in the Delphi programming language was used.

their rates of return have a certain internal structure, they can be chaotic. The exception are the time series CDPROJECT and WIG20.

According to Taqqu and Teverovsky [13] suggestion in the estimation of the parameter H we chose value of m parameter as integer part of $\frac{N}{32}$, where N denotes length of the considered time series. Estimation results are shown in the Table 2:

Company / Index	Estimated H value	Standard deviation of estimator	Test statistic	p-value	Market cap (mln zł)
SWIG80	**0.670**	0.0493	3.4535	0.001***	na
RELPOL	**0.661**	0.0493	3.2761	0.001***	76
INVESTORMS	**0.646**	0.0493	2.9697	0.003***	na
AMICA	**0.637**	0.0493	2.7786	0.005***	1 003
POLNORD	**0.625**	0.0493	2.5459	0.011**	273
MWIG40	**0.617**	0.0493	2.3806	0.017**	na
GROCLIN	**0.616**	0.0493	2.3469	0.019**	49
INGBSK	**0.611**	0.0493	2.2544	0.024**	26 150
KGHM	**0.595**	0.0493	1.9276	0.054*	18 460
FORTE	**0.591**	0.0493	1.8384	0.066*	1 282
DEBICA	**0.577**	0.0493	1.5626	0.118	1 705
CCC	**0.553**	0.0493	1.0673	0.286	11 138
WIG	**0.551**	0.0493	1.0288	0.304	na
CDPROJEKT	**0.549**	0.0493	0.9920	0.321	11 736
BZWBK	**0.525**	0.0493	0.4993	0.618	37 051
AGORA	**0.513**	0.0493	0.2684	0.788	667
PKOBP	**0.513**	0.0493	0.2587	0.796	51 562
WIG20	**0.503**	0.0493	0.0675	0.946	na
PKNORLEN	**0.471**	0.0493	-0.5856	0.558	38 921
PEKAO	**0.430**	0.0493	-1.4300	0.153	32 415

Table 2 Local Whittle Estimator: estimation results

First column contains the name of the financial instrument, while the second one presents the estimated value of parameter H for $m = 103$. Third column shows the estimated standard deviation of the estimator, while fourth column contains the test statistic with formula:

$$\frac{(\hat{H}-H)}{\sqrt{\frac{1}{4m}}}, \tag{18}$$

where as the testing value of parameter H we took $H = 0.5$, characteristic for white noise. Fifth column shows the p-value for two-tailed test, while sixth contains the market capitalization of selected companies as on 13.04.2018.

Results of estimation suggest, that in considered period instruments listed on the Warsaw Stock Exchange was characterized by a heterogenous parameter of long memory. For three indexes and seven companies both estimated values of H were greater than 0.5 and results was statistically significant. The highest value of H, equal to about 0.67, was estimated for SWIG80 index, that is computed from the share prices of small companies. This result may suggests, that in the area of small and medium-sized enterprises market isn't as efficient as in the case of larger companies. This dependence may be confirmed by the fact, that estimated value of H for index WIG20, which includes shares of the largest companies, wasn't significantly different from 0.5. Interestingly, Pearson's product-moment correlation coefficient between estimated long memory parameter H and companies capitalization is distingly negative and was computed on the level -0.66.

The results provides evidence of presence of long-term memory in stock returns for some financial instruments listed on Warsaw Stock Exchange. Values of H parameter significantly greater than 0.5 was mainly found for shares and indexes of small and medium-sized enterprises. However, the strength of such dependence isn't as significant as in case of some naturally occuring time series. One of the phenomenon studying most frequently in this context is the series of annual minimum water level of the River Nile, measured during the years 622

through 1284. Robinson [11] estimated H value for this series between 0.866 and 0.941, depending on the parameter m. This results differs significantly in plus from the estimates obtained in this paper for stock returns.

4 Conclusions

The results obtained for both estimators are relatively compatible. The only significant difference occurs in the case of a series of return rates of RELPOL company; another is the fact that the Hurst exponent does not indicate the presence of a slight anti-persitence for the PKN ORLEN and PEKAO series, which was detected by the Whittle estimator (although this was not a statistically significant result).

The question of long-term memory in financial time series occur most commonly in the context of efficient-market hypothesis, especially in its *weak* variant. In the next step we should try to find out, if for selected stock assets with the highest estimates of H parameters is it possible to achieve above-average profits using appropriate time series models.

References

[1] Chun, S.H., Kim, K.J., Kim, S.H.(2002). Chaotic analysis of predictability versus knowledge discovery techniques: case study of the Polish stock market. *Expert Systems*, *19*(5), 264 – 272.

[2] Czekaj, J., Woś, M. i Żarnowski, J. (2001). *Efektywność giełdowego rynku akcji w Polsce z perspektywy dziesięciolecia.* Warszawa: Wydawnictwo Naukowe PWN.

[3] Gurgul, H. & Wójtowicz, T. (2006). Długookresowe własności wolumenu obrotów i zmienności cen akcji na przykładzie spółek z indeksu DIJA. *Badania operacyjne i decyzyjne*, *3-4*, 29–56.

[4] Gurgul, H. & Wójtowicz, T. (2013). Zależności długookresowe zmienności stóp zwrotu i wielkości obrotów na GPW w Warszawie. *Innowacje w finansach i ubezpieczeniach - metody matematyczne i ekonometryczne*, *154*, 144–151.

[5] Gurgul H., Kłęk K., Syrek R. (2009). Długookresowe własności kursów walutowych – podwójna długa pamięć. *Ekonomia Menedżerska*, *5*, 63–79.

[6] Hurst, H. E. (1951). Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, *116*, 770–799.

[7] Kwiatkowski, J. (2017). *Procesy z długą pamięcią i modele ARFIMA*. www.home.umk.pl/~jkwiat/ARFIMA.pdf (access: 13.07.2017).

[8] Orzeszko, W. (2005). *Identyfikacja i prognozowanie chaosu deterministycznego w ekonomicznych szeregach czasowych*. Warszawa: Polskie Towarzystwo Ekonomiczne.

[9] Palma, W. (2007). *Long-memory time series.* Hoboken, New Jersey: John Wiley & Sons Inc.

[10] Peters, E. E. (1996). *Chaos and order in the capital markets.*Canada: John Wiley & Sons Inc.

[11] Robinson, P. (1995). Gaussian Semiparametric Estimation of long range dependence. *The annals of statistic*, *23*(5), 1630–1661.

[12] Stawicki J., Janiak E.A., Müller-Frączek I. (1997). Różnicowanie fraktalne szeregów czasowych – wykładnik Hursta i wymiar fraktalny. *Dynamiczne modele ekonometryczne*: materiały na V Ogólnopolskie Seminarium Naukowe, 9-11 września 1997, Towarzystwo Naukowe Organizacji i Kierownictwa „Dom Organizatora", Toruń, 35– 41.

[13] Taqqu, M. S. i Teverovsky, V. (1997). *Robustness of Whittle-type Estimators for Time Series with Long-Range Dependence*. https://pdfs.semanticscholar.org/d4cb/4c4955245f213a677bce7269d47f4b8db665.pdf (access: 14.07.2017).

[14] Zeug-Żebro, K., Dębicka, J., Kuśmierczyk, P., Łyko, J. (2013). *Wybrane modele matematyczne ekonomii. Decyzje i wybory*. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.

Bin packing and scheduling with due dates

Frantisek Koblasa[1], Miroslav Vavroušek[2]

Abstract. This paper is presenting combination of classical single machine batch scheduling problem with capacity constraint. Unlike most cases, where capacity problem is considered as one dimensional packing or knapsack problem, in this paper processor capacity is represented by 3D bin packing problem.
This research is inspired by real world problems, as scheduling of the heat-treatment furnace, where it is necessary to schedule workflow of machines. That usually consist of heterogeneous job batches and it is necessary to take in account its work space while also considering technological constraints.
There is proposed early research done by reviewing problems as single machine batch scheduling, Bin packing problem and scheduling problem with objective function of lateness. There are also suggested theoretical models based on the practical example of heat treating operations which are describing combination of before mentioned packing and scheduling problems. Constructive algorithm and random key based evolution algorithm are then tested on these problems to discuss its practical applications.

Keywords: Batching problem, Scheduling problem, Bin packing problem, Evolution algorithm, Due dates.

JEL Classification: C27
AMS Classification: L23

1 Introduction

Planning and scheduling is becoming more and more complicated as products customization level is rapidly increasing up to personifications. That leads not only to development and practical use of artificial intelligence usually represented by evolution algorithms [14], but also to adjusting basic constructive algorithms to describe real constraints as accurately as possible [10]. The main goal (objective function) is to get our product to the customer in a desirable time (due date).

This paper consists of three main chapters. The second chapter is describing basic models used for following modifications with goal to get close to real world problems. Following chapter defines a generation of a theoretical model focused on setting of an appropriate due date and to enrich classical Bin packing problem as well.. The fourth chapter is describing constructive and evolution algorithm together with results comparison.

2 Bin packing and scheduling

Bin packing problem (BPP) is one of the space decompositions problems like knapsack, cutting, container loading, stock, trim loss and strip problem and is considered as NP hard.

In bin packing problem there are given *n* rectangular items of different dimensions and bins each of usually (with few exception [7]) same capacity *c*. The goal is to assign each item to a bin such that number of total used bins is minimized. It may be assumed that all items have weights smaller than bin capacity. There is extensive research in bin packing problem, but very few are focusing on real-world problem conditions.

There is variant of the two-dimensional bin packing problem where each rectangle is assigned a due date and each bin has a fixed processing time. The objective is to minimize the number of bins and to minimize the maximum lateness of the rectangles. This problem is motivated by the cutting of stock sheets and the potential increased efficiency that might be gained by drawing on a larger pool of demand pieces by mixing orders, while also aiming to ensure a certain level of customer service. [2]

Bin Packing/Covering with Delivery [1] where, unlike classical packing, there is added condition where bin has to be covered so utilization of bin is at least 100%. Problem had also online setting. The items arrive one by

[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.vavroušek@tul.cz.

one and an item needs to be packed immediately when it is revealed. The objective function is to maximize profit generated by covering bin in due date. This research also discussed special two priorities of due date problem - when to decide to "close" bin to be processed and maximal number of simultaneously opened bins. However, due date is represented by – deliver just now, assuming that more bins are open, more time it will take to pack.

Not only tardiness (lateness), but earliness is considered in [12]. This research is addressing bin packing problem together with pull oriented single machine scheduling problem. Processing time setting is based on length of cut and time per distance. Due date is then defined by mean deviation of processing time or uniform random distribution with maximal due date as double of processing time. Objective is to minimize the weighted earliness-tardiness penalties. The follow up research [13] is considering same setup of due dates as [2]

Application of bin packing and scheduling can be found in medical [3] applications. There is a problem addressed, when due date is bounded with opening vial (release date) and with expiration date (maximal processing time). Problem is then solved by branch and bound algorithm.

The following chapter is focusing to model problems close to machine working space limitation together with due date constraint.

3 Design of theoretical models for 3D bin packing problem with due dates

Combination of 3D bin packing problem with due date condition is based on real world problem where bins are representing processor which its processing batch is constrained by working space. There are lots of applications in real world problems beginning with mechanical engineering [17] ending with progressive technologies of rapid prototyping or better said additive technologies [6, 19]. Unlike rapid prototyping, where manufacturing batch is classified as *S*-batch where processing time is equal to sum of each task in batch, this research is focused on *P*-batch class where batch processing time is equal to the maximum processing time among all tasks.

The main problem is to set processing time base on technology which is representing. This research is based on manufacturing technology of heat treating furnaces where processing time (1) of each part is based on material and dimensions of heated objects.

$$t_{Ai} = t_{teor} \cdot k_1 \cdot k_2 \cdot k_3 \cdot k_4 \tag{1}$$

Where (*t* is time and *k* is corrective coefficient):

- t_{Ai} is processing time of one item (time of heat treatment).
- t_{teor} is theoretical time of heat treatment (2) according material and size of characteristic dimension [4].

$$t_{teor} = t_p + t_s + t_t \tag{2}$$

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.
- k_2 is influence of heat treating furnace
- k_3 is influence of placement (gaps) in furnace.
- k_4 is division of recommended annealing temperature and recommended quenching temperature [15].

Then t_{bj} (3) is processing time of *j* batch (bin) full of *m* items

$$t_{Bj} = \max\{t_{A1}; t_{Ai}; \ldots; t_{Am}\} \tag{3}$$

The next step is to set a system how to define due dates for each packed object. The usual systems [2] are based on uniform processing time of each bin together with defining objective function according to C_{max} problem where is the goal to minimize total competition time. C_j is then competition time of each bin full of Q_b set of rectangle items

$$C_j = t_{Bj} b \tag{4}$$

Where *b* is corresponding sequence number of bin for $b=1,\ldots,B$. Then bin due date to meet (5) is

$$\sigma_b = \min dD_j \; ; \; j \in Q_j \tag{5}$$

Due date dD_j (8) is uniformly random number based on lower bound of number of bins [2] L_{bin} (7) where V_j (6)is volume of bin

$$V_j = W_j \times H_j \times D_j \tag{6}$$

$$L_{bin} = \left\lceil \frac{\sum_{j=1}^{n} w_i h_i d_i}{V_j} \right\rceil \tag{7}$$

$\beta \in \{0.8, 0.9, 0.1\}$ is coefficient influencing difficulty of a problem (lower coefficient is , harder it should be to solve problem with objective function to meet due date).

This is however based on two conditions which are not reflecting constraints of our problem - batch processing time t_b is always same and requirement of due date is for one batch the same.

That is why definition of dD_j (11) in this research is based on different processing times as in [12]

$$dD_i \geq t_{Ai} \tag{8}$$

$$t_{A\max} \geq C_j \tag{9}$$

$$t_{A\max} = \max t_{Ai} \tag{10}$$

$$dD_i = \left\lceil t_{Ai}, \beta t_{A\max} L_{bin} \left(1 + \frac{t_{Ai}}{t_{Ai} + t_{A\max}} \right) \right\rceil \tag{11}$$

Objective function is reflecting not only due date of processor (4), but of each part.

$$f(x) = \sum_{j=1}^{m} \sum_{i=1}^{n} \max\left(0, dC_j - dD_i\right) \tag{12}$$

$$dC_{j+1} = dC_j + C_{j+1} \tag{13}$$

Space decomposition is modelled base on previously used modified 3D BPP class of instances which are based on [11]. Instance has bin size W = H = D = 100 and five types of items with two types of material (uniformly random selected) are considered with (u.r. stands for „Uniformly random“):

- 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].

Each type is tested with due date defined by (12) with three difficulty classes β. For each type (1-5) there where generated 10 tests with n parts n ∈ <10; 20; 30, 40> so 1200 experiments total.

4 Algorithms and results

The following chapters are focussing on testing previously defined bin scheduling models and testing them on constructive and evolution algorithm.

4.1 Constructive algorithm

Constructive algorithm (CA) is based on previously presented approach for 2D and 3D bin packing problems [8].The set of placed boxes is arranged according to the due date *dD* (11) from the earliest to the latest required completion time. The rest of the algorithm follow usual pattern. 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. That means Local search is applied on the non fitting item. 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.

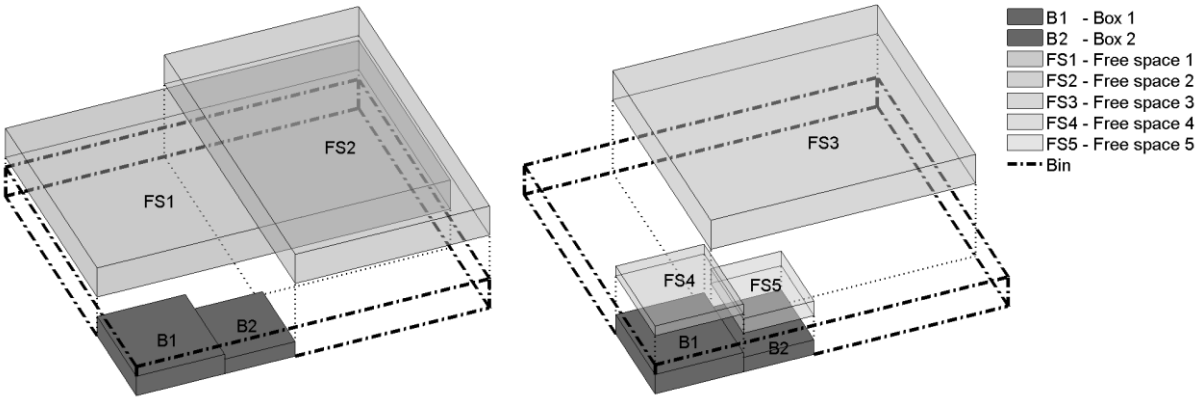


Figure 1: Box allocation in free space

4.2 Random key Evolution Algorithm

There is used random key Evolution Algorithm (EA) which shoved us previously good results in combinatorial problems [8,9].

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 then 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)*(12) by constructing solutions (see chapter 4.1)

3. Sort solutions base on their fitness function in descending order.

4. Select parents – in this case our selecting procedure differs from usual approach which follows basic random selection. So called biased principle is selecting parents from two different classes. First class called TOP is represented by percentage of best solutions in population and BOT in same way from the worst solutions.

This approach was highlighted [5] for its ability to keep heterogeneous population, so population does not converge that fast to local optimum. Number of selected parents is equal to size of population *N*

5. Crossover operation to make new offsprings by uniform crossover [16] with given probability.

6. Calculate fitness function *f(x)* (12) 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.

Evolution algorithm has that following setting:

- Population size *N*= 20**n* (where *n* is size of the problem.)
- Number of generations = 30
- Crossover probability = 0,7
- Crossover type Uniform
- Parent selection random, one from TOP 0,2, second from 0,2 BOT.

4.3 Computation results

Table of the experiment results (Fig 2) is showing percentage of objective function given by EA compared to the results given by CA. Results <0,1> marked light green (as 0.01) indicates percentage of EA lateness (eg 10% of CA value). Values marked by green and value of "0"means that EA was able to meet all due dates, when CA

was not. Grey field with value “1” means that both CA and EA find same solutions, however with delays. Yellow marking with value “-2” indicates CA and EA finds optimal solutions. Finally, red marked fields with value > 1 represents solutions where CA was able to get better results than EA.

	n	10					20					30					40				
model	Type	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
β=0.8	1	0.119	0.449	0.482	0.310	0.528	0.226	0.000	0.394	0.488	0.748	0.418	0.851	0.156	0.687	0.270	0.117	0.135	0.269	0.796	0.011
	2	0.894	0.688	1.147	0.916	1.000	0.000	0.593	0.689	0.704	0.850	0.620	0.248	0.197	0.698	0.489	0.321	0.125	0.342	0.640	0.137
	3	0.387	0.137	0.467	0.257	0.773	0.000	0.372	0.000	0.623	0.362	0.000	0.407	0.000	0.450	0.455	0.225	0.110	0.101	0.575	0.000
	4	1.000	0.325	-2.000	0.803	0.254	0.370	0.123	0.250	0.112	0.235	0.000	0.397	0.069	0.800	0.922	-2.000	0.000	0.000	0.769	0.191
	5	0.349	-2.000	0.691	1.000	1.000	1.000	0.409	0.243	0.353	0.090	0.367	0.207	0.326	0.509	0.107	0.366	-2.000	0.425	0.706	0.531
	6	0.093	0.000	-2.000	0.696	-2.000	0.126	0.126	0.000	0.805	0.458	0.520	0.198	0.314	0.375	0.669	1.421	-2.000	0.078	0.657	0.538
	7	0.592	0.414	0.184	-2.000	0.491	0.439	0.000	0.480	0.953	0.218	0.784	0.253	0.323	0.674	0.800	0.814	0.294	0.389	0.709	0.132
	8	0.547	0.518	0.364	1.000	1.000	0.014	0.000	0.288	0.873	0.177	0.264	0.321	0.124	0.429	0.000	0.174	0.060	0.486	0.539	0.197
	9	0.000	0.000	0.433	1.000	1.000	0.000	0.418	0.000	0.836	0.303	0.337	0.353	0.032	0.418	0.000	-2.000	0.336	0.830	0.505	0.639
	10	0.000	0.336	0.448	0.829	0.217	0.847	0.363	-2.000	0.438	0.409	-2.000	0.447	0.404	0.797	0.303	0.000	0.331	0.325	0.728	0.263
β=0.9	1	0.195	0.107	0.159	0.316	0.319	0.004	0.195	0.000	0.173	0.760	0.113	0.036	0.129	0.023	0.104	0.000	0.000	0.010	0.078	0.079
	2	0.000	0.624	0.000	0.000	1.000	0.085	0.112	0.040	0.000	0.395	0.020	0.149	0.016	0.042	0.258	0.000	0.008	0.000	0.146	0.035
	3	0.253	0.049	0.209	0.000	0.727	0.060	0.081	0.020	0.147	0.051	0.001	0.080	0.000	0.376	0.182	0.014	0.004	0.000	0.006	0.049
	4	0.500	0.000	0.578	0.768	0.023	0.251	0.266	0.000	0.076	0.305	0.000	0.111	0.000	0.002	0.120	0.000	0.000	0.000	0.124	0.079
	5	0.000	1.000	0.144	0.027	1.000	0.000	0.207	0.001	0.303	0.000	0.015	0.000	0.017	0.176	0.126	0.003	0.000	0.011	0.383	0.021
	6	0.057	0.403	0.000	0.285	-2.000	0.084	0.030	0.013	0.094	0.144	0.032	0.005	0.011	0.194	0.065	0.000	0.000	0.012	0.075	0.030
	7	0.369	0.000	0.169	0.000	0.117	0.020	0.030	0.020	0.344	0.225	0.039	0.011	0.000	0.072	0.109	0.000	0.017	0.021	0.148	0.091
	8	0.151	0.497	0.313	0.176	1.000	0.000	0.024	0.088	0.224	0.000	0.002	0.000	0.000	0.069	0.013	0.000	0.030	0.057	0.035	0.068
	9	0.002	0.581	0.000	0.486	1.000	0.000	0.001	0.000	0.109	0.000	0.009	0.147	0.093	0.105	0.012	0.000	0.001	0.000	0.093	0.060
	10	0.134	0.183	0.410	0.277	0.332	0.000	0.000	0.000	0.400	0.069	0.000	0.059	0.062	0.000	0.036	0.002	0.010	0.000	0.163	0.000
β=1	1	0.000	0.275	0.184	0.190	0.279	0.028	0.004	0.002	0.271	0.670	0.002	0.088	0.000	0.094	0.066	0.005	0.042	0.017	0.112	0.000
	2	0.000	-2.000	0.000	0.000	0.689	0.022	0.000	0.000	0.000	0.000	0.124	0.013	0.007	0.001	0.000	0.031	0.045	0.000	0.113	0.009
	3	0.000	0.000	0.287	0.215	0.763	0.034	0.273	0.167	0.056	0.013	0.000	0.100	0.000	0.025	0.020	0.000	0.067	0.000	0.001	0.105
	4	0.148	0.210	0.404	0.341	0.190	0.113	0.071	0.138	0.069	0.035	0.028	0.005	0.003	0.000	0.220	0.000	0.000	0.000	0.129	0.010
	5	0.000	0.392	0.000	0.389	1.000	0.000	0.077	0.000	0.041	0.085	0.000	0.000	0.252	0.005	0.005	0.004	0.000	0.000	0.044	0.106
	6	0.000	0.387	0.000	0.208	0.000	0.000	0.076	0.061	0.053	0.334	0.055	0.004	0.093	0.082	0.166	0.005	0.055	0.000	0.000	0.060
	7	0.320	0.499	0.220	0.000	0.003	0.068	0.000	0.000	0.187	0.189	0.000	0.028	0.000	0.010	0.054	0.050	0.005	0.000	0.056	0.113
	8	0.000	0.195	0.225	0.430	1.000	0.000	0.210	0.090	0.097	0.361	0.000	0.014	0.014	0.008	0.134	0.011	0.000	0.004	0.043	0.120
	9	0.223	1.000	0.146	0.283	1.000	0.000	0.000	0.000	0.001	0.116	0.000	0.000	0.002	0.020	0.075	0.000	0.102	0.077	0.046	0.080
	10	0.000	0.056	0.378	0.255	0.265	0.060	0.166	0.264	0.085	0.002	0.006	0.000	0.008	0.013	0.000	0.000	0.025	0.002	0.072	0.013

Figure 2: Objective function comparison

Analysing results, EA was in the 91.5% cases better than CA and in the 30% cases was able to find optimal solution. EA and CA found in 7.75% of experiments same solutions which in 3.25% cases were without delays. CA has found better solution only in two cases which makes 0.5% of the solutions.

The important aspect to consider is also to compare timespan of constructing solution by CA and using optimization by EA. If optimization takes longer that actual benefit, developed EA is than suitable primarily for offline solutions. CA has computation time less than second while EA, in the case of problem with 10 parts, around 40 seconds up to 40 minutes in the case of the biggest problem with 40 pieces (where 24 000 solutions were calculated during optimization). Comparing the minimal reduction of due date achieved (n= 10 reduction by 2000s, n=20 by 14 000s, n= 30 by 10 000s, n=40 by 20 000s) to optimization timespan, algorithm is adequate to be used in practice taking in the account this aspect.

Analysing showed results from combinatorial optimization point of view it is necessary to point out some interesting findings. The most of the cases (excluding one) when both CA and EA found same solution with some delays located in the field of the problem with 10 parts. This leads to conclusion that those problems are not very complex. The same situation is with the case when both CA and EA found solutions without delays were β=0.8.

5 Conclusion

The main contribution of this paper is to define theoretical models which will not follow just competition time, but much more needed real problems , where objective function is to meet customer demand in time – due date. Comparing to other approaches, paper added complexity of variable processing time based on the size of processed item. The presented development of bin packing models with due dates showed necessity to investigate its properties more in deep to find theoretical models with more difficult local optimums. The major reason is that in 7.75% of cases developed CA for due dates problem and EA found same results, which means those models were not that much complex to be solved by EA. That might be caused by number of scheduled components, their sizes and by system of setting due dates.

However, developed evolution algorithm showed very good results and the only question is if it is caused by local search mechanism responsible for rotating parts or evolution principles of EA. Local search principle will be further developed and rotation of parts will be included directly in to the chromosome in EA.

Future research will be focussed also on the combination of BPP and job shop scheduling.

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] Benkő, A., Dósa, G., & Tuza, Z. (2010). Bin Packing/Covering with Delivery, solved with the evolution of algorithms. In *Bio-Inspired Computing: Theories and Applications (BIC-TA), 2010 IEEE Fifth International Conference on* (pp. 298-302). IEEE.

[2] Bennell, J. A., Lee, L. S., & Potts, C. N. (2013). A genetic algorithm for two-dimensional bin packing with due dates. *International Journal of Production Economics*, *145*(2), 547-560.

[3] Billaut, J. C., D. Croce , & F. GROSSO (2013), A single machine scheduling problem with bin packing constraints. In 14th International Conference on Project Management and Scheduling (p. 40).

[4] Bryson, W. E.: *Heat Treatment Selection and Application of Tool Steels 2E*. Hanser Verlag, Munich, (2005), ISBN 978-3-446-41941-4

[5] Gonçalves, J. F., & Resende, M. G. (2013). A biased random key genetic algorithm for 2D and 3D bin packing problems. *International Journal of Production Economics*, *145*(2), 500-510.

[6] Griffiths, V., Scanlan, J. P., Eres, M. H., Martinez Sykora, A., & Chinchapatnam, P. (2018). Cost-driven build orientation and bin packing of parts in Selective Laser Melting (SLM). *European Journal of Operational Research*.

[7] Hong, S., Zhang, D., Lau, H. C., Zeng, X., & Si, Y. W. (2014). A hybrid heuristic algorithm for the 2D variable-sized bin packing problem. *European Journal of Operational Research*, *238*(1), 95-103.

[8] Koblasa, F., Vavroušek, M., & Manlig, F. (2015) Two-dimensional Bin Packing Problem in batch scheduling. In *Proceedings of 33rd International Conference Mathematical Methods in Economics* (Vol. 1, pp. 354-359).

[9] Koblasa, F. & F. Manlig (2014). Application of Adaptive Evolution Algorithm on real-world Flexible Job Shop Scheduling Problems. *In: 32nd International Conference MME2014*. First edition. Olomouc, s. 425-430. ISBN 978-80-244-4209-9.

[10] Lu, C., Gao, L., Li, X., & Xiao, S. (2017). A hybrid multi-objective grey wolf optimizer for dynamic scheduling in a real-world welding industry. *Engineering Applications of Artificial Intelligence*, *57*, 61-79.

[11] Martello, S., Pisinger, D., and Vigo, D (2000).: The three-dimensional bin packing problem. Operations Research, vol 48, 2, , 256-267.

[12] Polyakovskiy, S., Makarowsky, A., & M'Hallah, R. (2017). Just-in-time batch scheduling problem with two-dimensional bin packing constraints. In *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 321-328). ACM.

[13] Polyakovskiy, S., & M'Hallah, R. (2018). A hybrid feasibility constraints-guided search to the two-dimensional bin packing problem with due dates. *European Journal of Operational Research*, *266*(3), 819-839.

[14] Raska, P., & Ulrych, Z.. . (2014) Testing optimization methods on discrete event simulation models and testing functions. *Procedia Engineering*, 69, 768-777

[15] Smóling K. 1988. : *Tepelné a chemicko-tepelné spracovanie v príkladoch*. Alfa, Bratislava,

[16] Umbarkar, A. J., & Sheth, P. D. (2015). Crossover operators in genetic algorithms: A review. *ICTACT journal on soft computing*, *6*(1).

[17] Xie, X., Li, Y., & Zheng, Y. (2012). Multiple crane scheduling in batch annealing process with no-delay constraints for machine unloading. In: *International Conference on Information and Automation (ICIA),*. 597-601

[18] Zhang, Y., Gupta, R. K., & Bernard, A. (2016). Two-dimensional placement optimization for multi-parts production in additive manufacturing. *Robotics and Computer-Integrated Manufacturing*, 38, 102-117.

[19] Zheng, J.-N.; Chien, Ch.-F.; Gen, M. (2015): Multi-objective multi-population biased random-key genetic algorithm for the 3-D container loading problem. *Computers & Industrial Engineering*, 89, 80-87

Heuristic approach to solve various types of the zone tariff design problem

Michal Kohání[1]

Abstract. An effective integrated transport system (ITS) is one of ways how to motivate citizens to use public transportation. One of many sub-problems to be solved is the designing of tariff system. Several approaches to design of the tariff system exist. One of them is based on a division the region into smaller areas - tariff zones. When we want to design zone tariff system, we have to decide about optimal price for travelling and optimal zone partitioning of the region. The zone tariff can be divided into two types. The first one is the zone tariff with arbitrary prices, where fares are given for all pairs of zones arbitrary. Second type, the counting zone tariff system, is based on the calculation of travelled zones on the trip and the price of trip is based on the number of travelled zones. In this paper we describe both types of zone tariff system and we propose heuristic approach to solve them. The model of zone partitioning can be described with the mathematical model which is similar to the p-median problem. Mathematical models contain non-linear objective function what can be the main drawback of this approach. We propose heuristic approach based on a greedy heuristic method. We make a computational study on test data sets from selected region to compare both types of zone tariff systems.

Keywords: tariff zones design, zone partitioning, p-median problem, heuristic method.

JEL Classification: C44, C61, O18
AMS Classification: 90B06, 90B80, 90C10, 90C59

1 Introduction

Sustainable transportation and transport integration are connected with a large numbers of challenges, especially in public transportation within the cities and suburban regions. One of the way how to motivate citizens to use public transportation is to create Integrated Transportation Systems (ITS), that combine various transport modes and enables to passengers to use one ticket to all modes. Designing of tariff system in ITS is also an important issue, tariff system should be understandable and motivating for passengers, but, on the other hand, effective for transport operators and municipalities. [8] [10] [12]

Topic of the tariff design can be found in the literature only rarely. Hamacher and Schöbel in [3] and [4] and Schöbel in [10] described a basic division of tariff systems. First and frequently used approach is a distance tariff system, where the price for a trip depends only on the length of the trip. This tariff system can be considered as fair. The price for the trip is based on the distance between arrival and destination station of the trip. Second type of tariff is a unit tariff. All trips in this system cost the same price and are independent on their length.

Third type of tariff, which combines the unit tariff and the distance tariff, is a zone tariff system. In this system the whole area has to be divided into smaller sub-regions - tariff zones. The price for a trip in the zone tariff system is calculated using the information about the starting and the ending zone of the trip. The calculation of the price for the trip can be done in two different ways. In the first case, in a zone tariff with arbitrary prices (ZTAP), the fare calculation is based on the zone fare matrix, where the price is given arbitrarily for each pair of zones. Second possibility of the zone tariff system is a counting zone tariff system (CZTS). The price for travelling depends on the origin and destination zone and on the number of travelled zones on the trip. It holds that trips passing with various origin and destination zone but with the same number of crossed zones have the same price.

Zone Tariff Design Problem is usually connected with various decisions that have to be made. Since we need to create tariff zones and divide the region into the smaller regions, we have to decide about the number of tariff zones and about assigning of the stop/station in the particular tariff zone. This part of the problem can be called Zone Partitioning Problem. With these decisions is also connected the decision about the new tariff, what is in the literature called as Fare Problem.

[1] University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, Žilina, Slovakia, michal.kohani@fri.uniza.sk

In this contribution we describe both types of zone tariff system and we focus on the zone partitioning problem in both zone tariff types. In the Section 2 the model of zone partitioning is described by the mathematical model which is similar to the p-median problem. Mathematical models contain non-linearity what can be the main drawback of this approach. In the Section 3 we propose solution approach based on greedy heuristic method. In the Section 4 we make a computational study on test data sets from selected region to compare both types of zone tariff systems in terms of quality of the solution.

2 Mathematical models of the Zone Tariff Design Problem

Let all stations in the public transport network constitute the set of nodes I. Stations i and j from the set I are connected by the edge $(i,j) \in E$, if there is a direct connection by public transport line between these two stations. Symbol E denotes the set of edges. Distance between stations i and j is denoted as d_{ij}. For each pair of stations i and j is c_{ij} the current or fair price of travelling between these two stations.We assume that c_{ij} is equal to c_{ji} for each pair of stations i and j. n_{ij} represent the new price for travelling between i and j in the zone system. Number of passengers between stations i and j is b_{ij} (OD matrix). To describe passenger flows we introduce parameter a_{ij}^{rs}, where used paths will be observed. Value of the parameter a_{ij}^{rs} is equal to 1 if the edge (r,s) is used for travelling from station i to station j and 0 otherwise.

Construction of the zone partitioning model was inspired by the model of the p-median problem [5]. Based on this model, we introduce binary variables y_i, which represent a "fictional" centre of the zone and is equal to 1 if there is a centre of the zone in node i and 0 otherwise. For each pair of stations i and j we introduce variable z_{ij} which is equal to 1 if a station j is assigned to the zone with centre in the node i and 0 otherwise. We expect to create at most p_{max} tariff zones.

When we want to suggest the objective function of the model, there are many possible ways. In the papers by Hamacher and Schöbel [3] [4] were presented three possible objective functions. There are also other possibilities, such as maximization of transport operator revenue, which is also mentioned in [1] and [10]. In presented model we use the average weighted deviation between current and new price for all passengers, according to the advices of experts in [10] and previous research of exact approach to the problem in [5]. Objective function of the Zone Tariff Design Problem can be formulated as follows (1):

$$dev_{avg} = \frac{\sum_{i \in I} \sum_{j \in I} \left| c_{ij} - n_{ij} \right| b_{ij}}{\sum_{i \in I} \sum_{j \in I} b_{ij}} \tag{1}$$

Constraints that hold in the p-median problem are also necessary in the Zone Tariff Design Problem and can be formulated as follows:

$$\sum_{i \in I} z_{ij} = 1, \text{ for } j \in I \tag{2}$$

$$z_{ij} \leq y_i, \text{ for } i, j \in I \tag{3}$$

$$\sum_{i \in I} y_i \leq p_{\max} \tag{4}$$

$$z_{ij} \in \{0,1\}, \text{ for } i, j \in I \tag{5}$$

$$y_i \in \{0,1\}, \text{ for } i \in I \tag{6}$$

Conditions (2) ensure that each station will be assigned to one zone. Conditions (3) ensure that station j will be assigned only to the existing centre of the zone. Condition (4) ensures that we will create at most p tariff zones.

2.1 Mathematical Model of the Zone Tariff with Arbitrary Prices

As we mentioned before, the main difference between ZTAP and CZTS is in the price calculation. In the ZTAP the new price for travelling is calculated as the price between the zones where the origin and destination stops are assigned. To describe the zone fare matrix for calculating new price we introduce matrix h. Value h_{kl} describes the

price for travelling between the zone with zone centre in the node *k* and the zone with the centre in the node *l*. To calculate new prices for travelling between stops $i \in I$ and $j \in I$ we can use formula (7).

$$n_{ij} = h_{kl} z_{ki} z_{lj}, \text{ for } i, j, k, l \in I \tag{7}$$

Mathematical model of the Zone Tariff with Arbitrary Prices (ZTAP model) we can formulate as follows:

Minimize (1)

Subject to (2)-(6), (7)

2.2 Mathematical Model of the Counting Zones Tariff System

When we want to set a new price for travelling in Counting Zones Tariff System (CZTS), there are more possibilities how to do it. In [5] and [6] a unit price for travelling per one zone was set. In this paper we use two different unit prices, as was mentioned in [5] – price f_1 for travelling in the first zone and unit price f_2 for travelling in each additional zone. If we want to calculate new price of the trip between nodes *i* and *j* in the system, we need to calculate number of zones crossed on this trip. The calculation of the number of crossed zones can be easily replaced by the calculation of crossed zone borders accordingly to [4] and [10]. We assume that station can be assigned only to one zone and the border between zones is on the edge. We introduce binary variable w_{rs} for each existing edge $(r, s) \in E$, which is equal to 1 if stations *r* and *s* are in different zones and is equal to 0 otherwise. For the calculation of crossed borders number we need to use the path used for travelling between stations *i* and *j* described by parameters a_{ij}^{rs}. Determination of crossed zone borders is associated with following set of constraints:

$$z_{ij} - z_{ik} \le w_{jk}, \text{ for } i \in I, (j,k) \in E \tag{8}$$

$$w_{ij} \in \{0,1\}, \text{ for } (i, j) \in E \tag{9}$$

New price n_{ij} in CZTS can be determined by the number of crossed zones and calculated as follows (8):

$$n_{ij} = f_1 + \sum_{(r,s)\in E} f_2 a_{ij}^{rs} w_{rs}, \text{ for } i, j \in I \tag{10}$$

Mathemathical model of the Counting Zones Tariff System (CZTS model) we can formulate as follows:

Minimize (1)

Subject to (2)-(6), (8), (9), (10)

3 Heuristic method for Tariff Zones Design Problem

Objective function (1) in this model is not a linear function. Several solution approaches for designing zone tariff system can be found in the literature. Hamacher and Schöbel in [3] and [4], Schöbel in [10] and Babel and Kellerer in [1] proposed exact solution approaches for the counting zones tariff system where the goal was to design zones such that new and old price for most of the trips are as close as possible. A note on fair fare tariff on the bus line was mentioned also by Palúch in [9]. Another approach to solve this problem was proposed in [1], [4] and [10]. In the first stage optimal price of travelling is calculated and subsequently, three algorithms were used to calculate the zone partitioning. First algorithm is based on the clustering theory, second algorithm is a greedy algorithm and the last algorithm is based on the spanning tree approach. Heuristic approach based on greedy algorithm for zone merging with various parameters for zone size evaluation was mentioned also in [6] and [7].

3.1 Heuristic Approach for Zone Partitioning Problem

Based on the results of numerical experiments in [6] and [7] we use a greedy based heuristic method to solve the zone partitioning problem originally proposed in [6]. The algorithm starts with the situation, where each station forms separate zone and the number of zones is equal to the number of stations. The algorithm subsequently select two smallest neighboring zones which are merged into the one new zone. The goal is to create zones that are approximately the same size. Algorithm terminates when given maximum number of zones p is reached. Greedy heuristic algorithm with approximately same zone size can be described as follows:

STEP 1: Start with a partition *P* consisting of $|I|$ zones, each zone contains a single station. For each zone Z_i from *P* calculate the parameter e_i to express size of the zone.

STEP 2: Determine two neighboring zones Z_i and Z_j, where the sum $e_i + e_j$ is minimal.

STEP 3: Merge zones Z_i and Z_j into new zone Z_k and get a new partition P. For zone Z_k calculate parameter e_k.

STEP 4: If maximum number p_{max} of zones is reached, then terminate, else go to STEP 2.

In all formulas we use following notation: S_i is the set of stops, which are connected with at least one stop in the zone Z_i, $|Z_i|$ represents the number of nodes in the zone Z_i and parameter b_k represents the number of inhabitants in the node k.

For the calculation of zone size parameter e_i we can use various approaches. In [6] there were proposed various formulas and we select two formulas which give best solutions. Formula (11) calculates parameter e_i as the average distance of all stops in the zone Z_i to neighboring zones. Greedy heuristic with parameter (11) is denoted as H1.

$$e_i = \frac{\sum_{k \in Z_i} \sum_{j \in S_i} d_{kj}}{|Z_i| . |S_i|} \tag{11}$$

Formula (12) use number of inhabitants of all stations in the zone Z_i as the zone size parameter e_i. Greedy heuristic with parameter (12) is denoted as H2.

$$e_i = \sum_{j \in Z_i} b_j \tag{12}$$

3.2 Calculation of new prices

In the case of new prices calculation in TZAP, we have to determine price for all pairs of new zones (zone tariff matrix h). We were inspired by the new price calculation described in [3]. We propose 3 different ways of new price calculation. In formula (13), as the price for travelling between zones with centres in nodes k and l we select minimal price for travelling between all stops i from the zone Z_k and stops j from the zone Z_l.

$$h_{kl}^{\min} = \min\{c_{ij} : i \in Z_k, j \in Z_l, i, j \in I\} \tag{13}$$

In formula (14), as the price for travelling between zones with centres in nodes k and l we select maximal price for travelling between all stops i from the zone Z_k and stops j from the zone Z_l. In formula (15) we expect to calculate new price as the average between h_{kl}^{min} and h_{kl}^{max}.

$$h_{kl}^{\max} = \max\{c_{ij} : i \in Z_k, j \in Z_l, i, j \in I\} \tag{14}$$

$$h_{kl}^{avg} = \frac{h_{kl}^{\min} + h_{kl}^{\max}}{2} \quad for\ k, l \in I \tag{15}$$

To obtain the best settings of parameters f_1 and f_2 in CZTS we apply formula (10) with various setting of parameters f_1 and f_2 and we subsequently calculate average difference between new and old prices for all possible settings of parameters f_1 and f_2. As the best solution we choose the one that has the smallest value of the average difference.

4 Numerical Experiments

The goal of numerical experiments is to compare solutions obtained using TZAP model and CZTS model. We make the computational study on the data sets created from the real public transportation network in the Zvolen Region (Slovak Republic) in the year 2016. These data were obtained from [14]. Although the transportation network consist of large number of single stops, we group these stops by affiliation to the municipalities, what is commonly used technique and therefore we can also reduce the size of solved problem. We use two networks with 25 or 51 stations, which are shown in the Figure 1.

Current prices c_{ij} were calculated according to real prices depending on the shortest distance for travelling between all pairs of stops by regional buses. The OD matrix was estimated using the gravity model as in [2]. Numerical experiments were performed on the personal computer equipped with Intel Core 2 Duo E6850 with parameters 3 GHz and 3.5 GB RAM. Experiments were performed in Xpress Optimizer solver [11]. To solve test instances we use heuristic approaches H1 and H2 described in section 3. In the case of CZTS model, according to

the current fare prices we set the values of parameter f_1 from 0.3 to 0.9 with step by 0.05 and values of parameter f_2 from 0.1 to 0.6 with step by 0.05 for all the experiments and as the result we select the objective function with minimal value. Due to the grouping of stations, we set the nonzero price also for the trips with zero distance (travelling inside of the municipality) to avoid zero prices in the calculation of prices in TZAP model.

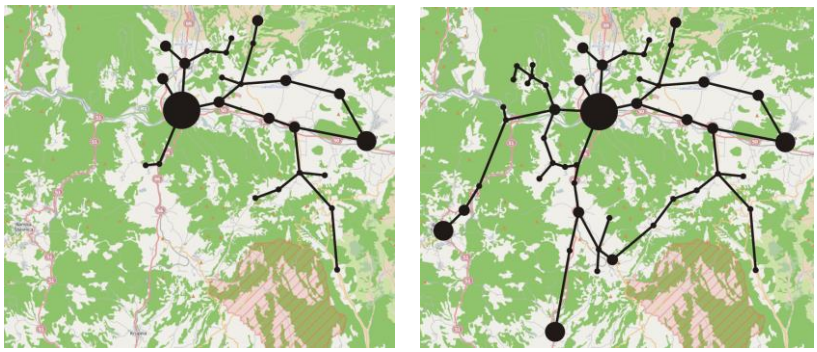


Figure 1 Test networks with 25 and 51 stations, map source: openstreetmap.org

In the Table 1 there are results for the test network with 25 nodes and in the Table 2 there are result for the network with 51 nodes. In all tables we denote columns as follows. Columns denoted as CZTS represent results of the CZTS model and columns denoted as ZTAP represent results for the TZAP model. Column denoted as H1 represents results for greedy heuristic H1 with the zone size calculation using formula (12) and column denoted as H2 represent result for heuristic method H2 with using of formula (13). Columns denoted as h^{min} in the case of TZAP model contain results for new zone tariff matrix calculation with formula (13), columns h^{max} contain results for new zone tariff matrix calculation with formula (14) and finally columns denoted as h^{avg} contain results for new zone tariff matrix calculation with formula (15).

p_{max}	*CZTS*		*ZTAP*					
	H1	*H2*	*H1*			*H2*		
			h^{min}	h^{max}	h^{avg}	h^{min}	h^{max}	h^{avg}
4	3367	4846	4273	25098	10769	4672	26184	11342
6	3525	4364	2564	17737	7943	2264	17672	7894
8	3374	2976	2300	10418	5057	1310	10170	4823
10	2793	2157	1230	9810	4883	714	6264	3097
13	2702	1751	1014	6445	3300	291	1179	615
16	2264	1351	203	563	379	134	344	237
20	1812	1371	74	178	124	55	125	90

Table 1 Best values of average deviation – networks with 25 stations

p_{max}	*CZTS*		*ZTAP*					
	H1	*H2*	*H1*			*H2*		
			h^{min}	h^{max}	h^{avg}	h^{min}	h^{max}	h^{avg}
4	8460	10629	11937	55855	23602	11956	47392	19815
6	8805	9348	10232	43817	18992	10406	30358	14737
8	7723	9168	9683	38653	16912	6040	27768	12959
10	7196	8027	8892	37400	16539	4268	20257	9307
13	6863	6755	5340	19222	9020	2824	17237	8028
16	6339	5906	3839	17429	8442	1943	11379	5404
20	5461	5264	3499	12916	6444	1313	7832	3863
25	4940	3849	1851	10243	5004	590	3552	1808
30	3595	3368	807	2657	1656	329	1047	610

Table 2 Best values of average deviation – networks with 51 stations

5 Conclusion

In the paper we presented various approaches to the Zone Tariff System Design. We presented models for Counting Zones Tariff System and Zone Tariff with Arbitrary Prices that are able to solve the zone partitioning and also calculate new prices for travelling in new zone system. We performed numerical experiments with both models using heuristic method that is based on the greedy approach and stepwise merging of neighboring zones. Numerical experiments were performed on the data sets created from the public transport network in Zvolen Region in Slovakia.

From the results of the numerical experiments in the Table 1 and Table 2 we can see the differences in obtained solutions using both models. In general, ZTAP model seems to give better results for the situation when the number of created zones is higher. On the contrary, CZTS model gives better results for situations, where the number of created zones is smaller. This information can be useful for system planners who design tariff systems and evaluate possible situations.

When we focus on the solution method – greedy heuristic, we can see, that behavior of heuristics H1 and H2 is different depending on used model and size of the problem. In most cases, heuristic method H2 gives better solution than heuristic H1, only when solving CZTS model with smaller parameter p, the situation is opposite. When we evaluate results of the TZAP model from the point of view of price calculation methods, we can see that the best results we can obtain applying the formula (13).

In the future we would like to verify this approach on the different networks with real OD matrix. Numerical experiments will be also extended to the larger test network to study behavior of presented models and heuristics on larger problems.

Acknowledgements

This work was supported by the research grants VEGA 1/0463/16 “Economically efficient charging infrastructure deployment for electric vehicles in smart cities and communities” and APVV 15-0179 “Reliability of emergency systems on infrastructure with uncertain functionality of critical elements”.

References

[1] Babel, L., Kellerer, H. (2003). Design of tariff zones in public transportation networks: theoretical results and heuristics, *Mathematical Methods in Operations Research* 58
[2] Cenek, P., Janáček, J. and Jánošíková, Ľ. (2002). Location of Transportation Districts at Modeling of Tranportation Processes in a Region. *Communications - Scientific Letters of the University of Zilina*, 4: 1-2.
[3] Hamacher, H., W., Schöbel, A. (1995). On Fair Zone Designs in Public Transporation, *Computer-Added Transport Scheduling*, 8-22.
[4] Hamacher, H., W., Schöbel, A. (2004). Design of Zone Tariff Systems in Public Transportation, *Operations Research* 52, 897-908.
[5] Kohání, M. (2013). Designing of Zone Tariff in Integrated Transport Systems, *Communications: scientific letters of the University of Žilina* 15, No. 1
[6] Kohání, M. (2014). Heuristic Approach to the Zone Partitioning Problem in Counting Zones Tariff System. *MME: proceedings of the 32th international conference. 2014*, Vol. 32: 449-454.
[7] Kohání, M. (2015). Various modifications of greedy algorithm for the zone partitioning problem in zone tariff systems. *SOR ´15: proceedings of the 13th International Symposium on Operational Research*, 221-226.
[8] Pratelli, A., (2004).The combined zone and fare planning problem, *Urban Transport X*, C. A. Brebbia & L. C. Wadhwa (Editors), WIT Press, 311-320.
[9] Palúch, S. (2013). On a fair fare rating on a bus line, *Communications: scientific letters of the University of Žilina* 15, 1
[10] Schöbel, A. (2006). *Optimization in Public Transportation: Stop Location, Delay Management and Tariff Zone Design in a Public Transportation Network*, Springer, 2006
[11] *XPRESS-Mosel User guide* (2012), Fair Isaac Corporation, Birmingham
[12] *BID* (2015). Bratislavská integrovaná doprava (Bratislava Integrated Transport Office). http://www.bid.sk [Accessed 5/5/2015]
[13] *OREDO* (2016). Integrated Transport System Organizator – Eastern Bohemia, http://www.oredo.cz/tarifni-mapy-iredo/ [Accessed 10/9/2016]
[14] Timetables of transport operator SAD Zvolen (2016), http://www.sadzv.sk/ [Accessed 11/9/2016]

The effects of fiscal policy shocks in Czech and German economy: SVAR model with graphical modeling approach

Eliška Komárková[1], Vlastimil Reichel[2]

Abstract. The paper will focus on identifying the influence of the Czech Republic (and Germany) fiscal policy on the business cycle of the Czech Republic (and Germany). The paper will state the fiscal policies of the states and their comparisons with the theory of real business cycle theory (RBC) and the New Keynesian theory (NK). A cascade of structural VAR models will be used to determine the impact of the fiscal policy shocks. The graphical modeling method will be used to determine the restrictions in the structural VAR models. The following variables will be used to estimate VAR models: government spending, government taxes, real output, consumption, working hours and real wages. The results obtained from the models will be used to assess the effectiveness of the fiscal policy (government spending shocks) and their impacts on the economic cycle. The results are in line with Keynesian theoretical models. Significant differences in impulse responses for individual economies can also be observed.

Keywords: Fiscal policy, Structural VAR model, Graphical Modelling.

JEL classification: E320, C130
AMS classification: 62M10, 65S05

1 Introduction

One of the most significant economic events that happened in the new millennium was the financial crisis of 2008/2009, which subsequently resulted in the economic crisis. This crisis was of a global nature and also affected most of the European Union (EU) states, including Germany and the Czech Republic. The article will focus on identifying the influence of the Czech Republic (and Germany) fiscal policy on the economic cycle of the Czech Republic (and Germany) during this period and will determine whether the empirical findings correspond with the theory of real economic cycles (RBC) or New Keynesian (NK).

Fiscal policies have a significant impact on the macroeconomic stability of countries. The government can use fiscal policy instruments to stabilize short-term fluctuations in the economy, whether it is a period of recession or boom. On the budget expenditure side, governments can make decisions about the level and composition of government spending. On the revenue side, they can make decisions about the amount of taxes.

In the 1980s, two economic lines emerged, a new classical school and a new Keynesian school. The representatives of the neoclassical school promoted the theory of real economic cycles based on assumptions of rational expectations and automatic purging of markets. On the contrary, the New Keynesian approach is based on the theory of monetary economic cycles. They do not question the theory of rational expectations, but they do not agree with clearing the markets in the short run. New Keynesians argue that nominal wages and prices are not so flexible in reality, which means that they cannot immediately adapt. Aggregate demand, therefore, does not adjust to aggregate supply and needs to be stimulated through fiscal intervention. The fundamental difference between these two theories is in the assumptions. While RBC's agents are subject to Ricardo's equivalence, in Keynesian Economic Cycle Theory, household consumption depends only on today's disposable income. According to Keynesians, macroeconomic variables respond differently to positive government shock. In response to the initial government spending shock, consumption, output, real wages and worked hours will increase. Specific reactions of selected macroeconomic variables to fiscal expansion are shown in Table 1.

	Output	Consumption	Worked hours	Real wage
RBC theory	+	-	+	-
NK theory	+	+	+	+

Table 1 Theoretical directions of macroeconomic variables responses to fiscal expansion

[1]Masaryk University, Lipová 41a, Brno, Czech Republic, 405874@mail.muni.cz
[2]Masaryk University, Lipová 41a, Brno, Czech Republic, reichel.v@mail.muni.cz

2 Data and methodology

2.1 Data

Time series of economies of the Czech Republic and Germany are used. The data were taken from the Eurostat statistical portals (Eurostat, [6]) and the OECD international organizations (OECD, [12]). The primary macroeconomic indicators are government spending (G_t), taxes (T_t) and real output (Y_t). Other variables will be gradually added to the model. Individually, private consumption (C_t), worked hours (H_t) and real wages (RW_t). All macroeconomic indicators are adjusted for inflation and converted per capita. Time series are quarterly, and their range of data, concerning availability, varies slightly for each state. The analysis is made on quarterly data of the Czech economy covering period 1999:Q1-2017:Q2 and quarterly data of German economy covering period 2002:Q1-2017:Q2. Most of the variables are seasonally adjusted from the source, only the real tax variable and the worked hours show seasonality. Seasonality was cleared using the X-12-ARIMA method.

	Variable	Gap	Source
Government spending	G	$\hat{g}$	Eurostat
Government taxes	T	$\hat{t}$	Eurostat
Gros domestic product	Y	$\hat{y}$	Eurostat
Consumption	C	$\hat{c}$	Eurostat
Worked hours	H	$\hat{h}$	OECD
Real wages	WT	$\hat{wt}$	Eurostat

Table 2 Time series description

2.2 Methodology

Structural VAR (SVAR) models have become a modern tool in analyzing the effects of monetary policy and sources of fluctuations in business cycles (Gottschalk [8]). The primary source of SVAR model when analyzing causes of fluctuation in the economic cycle is Blanchard and Quah ([1]). SVAR models were also used for their analysis Blanchard and Perotti [2] or Burnside, Eichenbaum and Fisher [3] or Fragetta and Melina [7]. The authors focus on detecting the effects of fiscal policy shocks to capture the impulse response of macroeconomic variables. Blanchard and Perotti [2] argues that SVAR methodology is more effective in analyzing the impacts of fiscal policy than of impacts of monetary policy. They state two major arguments. The first is that, in contrast to monetary policy, fiscal variables are moving for many reasons and are therefore considered to be exogenous fiscal shocks. The second argument is that, due to decision-making and implementation delays in fiscal policy, there is little or no discretionary fiscal policy response to unexpected movements in economic activity. Based on this, they can construct estimates of the automatic effects of fiscal variables on unexpected shifts in the economy to receive estimates of fiscal policy shocks. After identifying these shocks, they can monitor the dynamic effects of the gross domestic product and its components.

SVAR model

The VAR model is used for identification of fiscal policy shock. With variables of interest collected in the k-dimensional vector Y_t the reduced-form VAR model can be written as:

$$Y_t = \sum_{i=1}^{p} A_i \cdot Y_{t-i} + u_t, \tag{1}$$

where Y_{t-p} is vector of variables lagged by p periods, A_i is a time-invariant $k \times k$ matrix and u_t is a $k \times 1$ vector of error terms. As the error terms can be correlated, the reduced-form model is transformed into a structural model. Pre-multiplying both sides of equation 1 by the $(k \times k)$ matrix A_0, yields the structural form:

$$A_0 Y_t = A_0 A_1 \cdot Y_{t-1} + A_0 A_2 \cdot Y_{t-2} + \cdots + A_0 A_p \cdot Y_{t-p} + B\epsilon_t. \tag{2}$$

The relationship between structural disturbances ϵ_t and reduced-form disturbances u_t is described by equation of AB model (see Lutkepohl [10] or Lutkepohl and Kratzig [11]):

$$A_0 u_t = B\epsilon_t, \tag{3}$$

where A_0 also describes the contemporaneous relation among the endogenous variables and B is a $(k \times k)$ matrix. In the structural model, disturbances are assumed to be uncorrelated with each other. In other words, the covariance matrix of structural disturbances Σ_e is diagonal. The model described by equation (2) cannot be identified. Therefore, first the matrix B is restricted to $(k \times k)$ diagonal matrix. As a result, diagonal elements of matrix B represent estimated standard deviations of the structural shocks.

The next step is to impose restrictions to matrix A_0. The matrix A_0 is a lower triangular matrix with additional restrictions that are obtained from statistical properties of the data. The starting point of the process is a computation of partial correlations among the variables in the model and subsequent elimination of statistically insignificant relations. This methodology is based on well defined statistical rules and allows to derive many possible structural VAR models that are later evaluated by statistical information criteria.

Graphical modelling

Graphical modelling (GM) theory is a data-driven tool that allows to identify limitations directly from statistical data properties. It offers a systematic procedure to identify the SVAR model. In general, this methodology is dealt with by Darroch et al. [5]. Oxley [13] provides examples of how this method can be used in the macroeconomic analysis. Fragetta and Melina [7] uses this tool to analyze the effects of fiscal policy shocks. They designed the construction in several steps.

1. First, the partial correlation of residual components between variables will be calculated from the estimated VAR model.
2. Based on these partial correlations the Conditional Independence Graph (CIG), whose scheme is shown in Figure 1, will be constructed. Knots A, B, and C represent the variable or residual component of the VAR model. The edges joining the nodes represent partial correlations. Thus, CIG graphically represents all statistically significant interconnections or correlations between given variables.

Figure 1 CIG construction

3. Directed Acyclical Graphs (DAG) will be derived from the CIG graphs. Hypothetically, we can obtain 2^n DAG representations where the number 2 indicates the number of directions the edge can take and the n coefficient indicates number of these edges. For this illustrative example, there are four possible views of the DAG graphs that can be seen in Figure 2. DAG charts display the causal relationship of variables and represent all possible structures of SVAR models. Arrows between nodes indicate the direction of statistical causality.

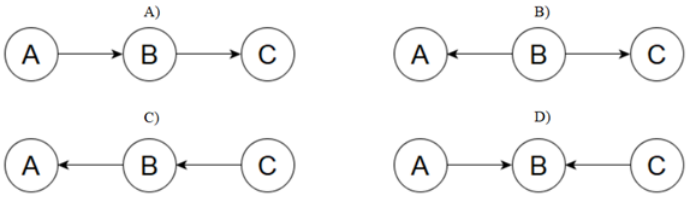


Figure 2 DAG charts - all possibilities

In deciding which of these DAG views will be the most accurate, we first apply the moralization rule. This rule states that if variables A and C determine B there is also a significant statistical correlation between A and C. Looking at case D), where A and C determine B, the corresponding CIG graph would have to be modified and nodes A and C would also have to be linked. By this definition, about the illustrative CIG, we exclude option D). For the analysis of the remaining DAG options, the t-statistics of the estimated regression coefficients of the VAR model will be used to determine the statistically significant causality between the variables. Causalities will be tested at a 5% significance level and the critical discipline will correspond to the interval $W \in (-\infty; -1,96) \cup (1,96; \infty)$.

3 Results

3.1 Results of CIG and DAG charts

From the estimates of partial correlations of residual components of the VAR model and application of GM theory rules, CIG graphs can be constructed and consequently DAG schemes derived. All types of schemes are captured

in Table 4. Schemes are key to determining the additional A_0 matrix restrictions mentioned in the equation (2). For example, from the DAG chart for the baseline VAR model (shown in Table 4), we can deduce the restrictions of the A_0 matrix which is mentioned in the table 3.

A_0 of baseline SVAR Germany

$$\begin{pmatrix} 1 & & \\ -a_{21} & 1 & \\ 0 & -a_{32} & 1 \end{pmatrix}$$

A_0 of baseline SVAR Czech Republic

$$\begin{pmatrix} 1 & & \\ 0 & 1 & \\ -a_{31} & -a_{32} & 1 \end{pmatrix}$$

Table 3 Identification of matrix A_0 in baseline SVAR models

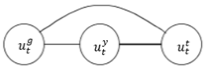
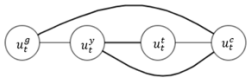
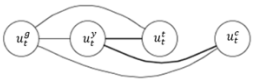
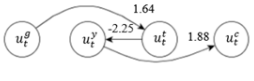
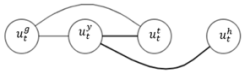
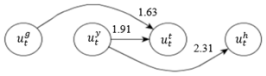
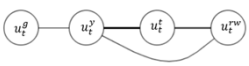
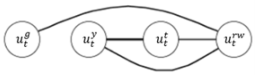
	Germany		Czech Republic	
	CIG	DAG	CIG	DAG
Baseline				
Consumption				
Working hours				
Real wage				

Table 4 Results of CIG and DAG charts

3.2 Impulse responses functions

Figure 3 and Figure 4 illustrate the impulse response of all selected macroeconomic variables to fiscal expansion (shock in government spending), along with percentile confidence interval calculated according to Hall [9]. The shock value is the size of one standard deviation. In the first graphs of Figure 3 and Figure 4, government spending responses are displayed, ie how the variable responds to its own shock. Government spending in both states rises first and return to its steady state after two quarters. It can be seen that shocks in both cases are not very persistent.

In the case of Germany (Figure 4), the impact of GDP falls below its steady level.The reaction is not statistically significant. More interesting is the later development of the impulse response of the output, which starts growing after the first quarter. In the fifth quarter reaches its maximum of 0.2%. Here we observe the statistical significance of the impulse response. It then slowly stabilizes to its steady state. The impulse response of taxes is very similar to the development of the impulse response of the gross domestic product. An analogous development of impulse responses for tax and GDP variables can be justified by the fact that tax collection depends on economic income.

For the Czech economy (Figure 3), the results of estimates for taxes and GDP are completely different from German. Output has a positive reaction immediately after the impact of the shock, but it immediately starts to decline and drops to -0.4 percent after six quarters. The reaction in minimum is statistically significant. Taxes also respond positively to the first impact of government spending and increase by about 1.1% after one quarter. Subsequently, they begin to decline, as it is in the case with GDP.

Consumption (in the case of Germany Figure 4) reacts negatively to government spending shock and then begins to grow. After a six quarters of shock, it rises to 0.1%. The impulse response is statistically significant when the response of this variable reaches a value slightly below 0.1%. Consequently, private consumption returns to its steady state. The worked hours fall by 0.05% in the first reaction of the positive government spending shock. After three quarters, it rises to 0.1%, and then the variable is stabilized to its steady state. The impulse response for the variable worked hours is statistically significant only after four quarters of the impact of fiscal expansion. Depending on the government's shock, employment rises as economic output has rises, just one period earlier.

Real wages respond after two quarters of the impact of fiscal shock by increasing to 0.2 percent. Then it also returns to its steady state.

The results of impulse responses for the Czech economy (Figure 3) are different from the impulse response of the German economy (Figure 4). Private consumption, as one of the components of Gross Domestic Product, is copying output development. At the impact of the shock, consumption increases and then falls below its steady state. After six quarters, it begins to grow, but will no longer exceed its steady state. The variable worked hours does not show any significant responses to government shock. For real wages, statistically significant estimates can be observed. In the impact, real wage rises to 0.8 percent, after the first quarter it will reach 1.2 percent. In the next development, it begins to fall to a steady state. Unfortunately, most confidence interval estimates for other variables contain zero and are therefore not statistically significant.

Compared with economic theories, the response of variables (for Germany): output, private consumption, worked hours, and real wages fully matches Keynesian theory. In response to the expansionary fiscal shock in government spending, these variables will increase. Impulse responses obtained from the Czech economy data are not in line with any of the above-mentioned theories.

The results of impulse responses are not surprising in the context of the events of the last decades. In response to the economic crisis, the German government has taken several measures aimed at mitigating the effects of the recession and restoring the German economy (Cupalová, [4]). The Czech government has not been confident in the crisis, which has negatively affected its ability to respond to the peak economic crisis (Vláda ČR, [14]).

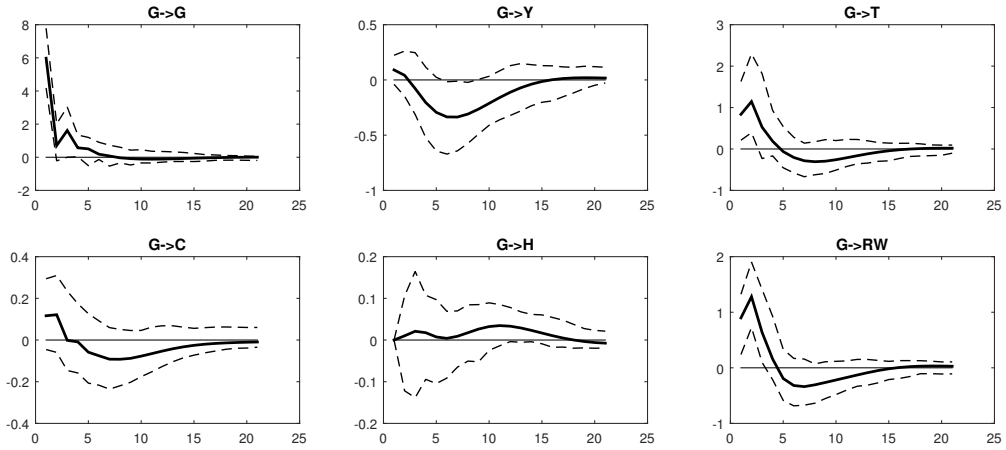


Figure 3 Czech Republic - Impulse responses functions to government spending shock

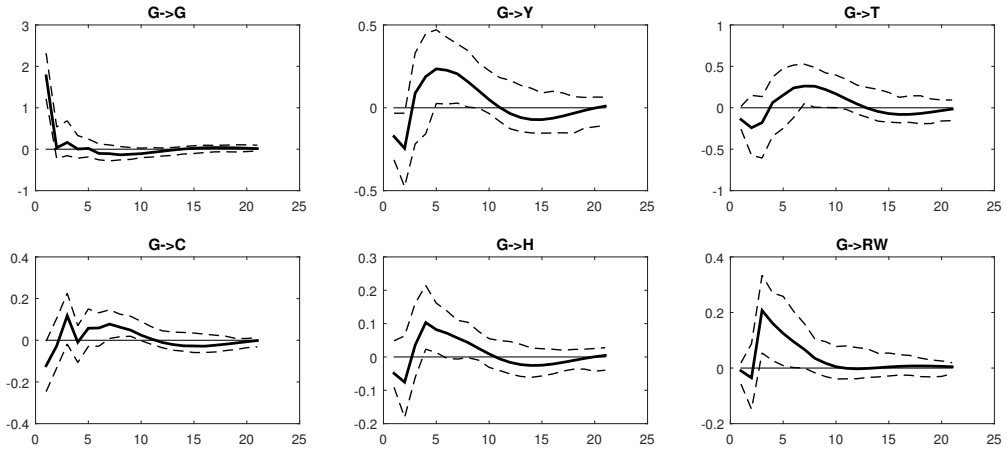


Figure 4 Germany - Impulse responses functions to government spending shock

4 Conclusion

In this paper, using the structural VAR model the impacts of fiscal expansion in the form of increased government spending were analyzed. To identify the structural component of the VAR model, we identified short-term restrictions derived from the graphical modeling approach were identified. Impacts of fiscal expansion were discussed using impulse response results. The results (captured in Table 5) show the difference between output and consumption responses in the Czech Republic and Germany. In Germany, there is an increase in consumption and an increase in output due to fiscal expansion. In the Czech Republic, on the other hand, the same shock results in a decrease in output and consumption. In the case of worked hours and real wages, we conclude that the variables respond to the government spending shock positively in both countries. Comparing the resulting impacts with the theoretical impacts, we see that Germany is consistent with the Neokeynesian theoretical concept of transmission. In the Czech Republic, it is unclear whether neokeynesian or RBC approaches can be used.

	Output	Consuption	Worked hours	Real wage
RBC theory	+	-	+	-
NK theory	+	+	+	+
Germany	+	+	+	+
Czech Republic	-	-	+	+

Table 5 Theoretical and observed directions of macroeconomic variables responses to fiscal expansion

Acknowledgements

This work was supported by funding of specific research at Faculty of Economics and Administration, Masaryk University, project MUNI/A/0966/2017. This support is gratefully acknowledged.

References

[1] Blanchard, O. J. & Quah D. T. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *The American Economic Review.* 79(4), 655673.

[2] Blanchard, O. & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics*, 117, 4, pp. 132968.

[3] Burnside, C., M. Eichenbaum, & Fisher J. D. (2004). Fiscal Shocks and Their Consequences. *Journal of Economic Theory*, 115, pp.89-117.

[4] Cupalová, M. (2009). Dopady světové ekonomické krize na veřejné rozpočty v zemích EU a vývoj českých veřejných financí. *Parlament ČR*.

[5] Darroch, J., Lauritzen, S. & Speed, T. (1980). Markov fields and loglinear interaction models for contingency tables. *The Annals of Statistics*, 8, pp. 52239.

[6] Eurostat. (2018). *Database Eurostat*. http://ec.europa.eu/eurostat/data/database

[7] Fragetta, M. & Melina.G. (2010). The Effects of Fiscal Shocks in SVAR Models: A Graphical Modelling Approach, *Birkbeck Working Papers in Economics and Finance 1006*. Department of Economics, Mathematics and Statistics, Birkbeck.

[8] Gottschalk, J. (2001). An Introduction into the SVAR Methodology: Identification, Interpretation and Limitations of SVAR models. *Kiel Institute of World Economics.*

[9] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

[10] Lutkepohl, H. (2006). *New introduction to multiple time series analysis*. Springer, Berlin.

[11] Lutkepohl, H. & Kratzig M. (2004). *Applied time series econometrics*. Cambridge University Press, Cambridge.

[12] OECD. (2018). *OECD database*. https://data.oecd.org/

[13] Oxley, L., Wilson, G. & Reale, M. (2009). Constructing structural VAR models with Conditional Independence Graphs. *Mathematics and Computers in Simulations*, 79, pp. 29166.

[14] Vláda ČR. (2009). *Národní protikrizový plán*.

World Tax Index: Results of the Pilot Project

Zlatica Konôpková[1], Jakub Buček[2]

Abstract. Taxes are the essential parts of models of economic growth. They may have negative impact on economy and burden part of the economic activity. Therefore, we need to be able measure the size of tax burden correspondingly. The most popular indices are tax quota and implicit tax rates, but there are also other alternative indicators including World Tax Index (WTI). WTI is an overall multi-criteria indicator of the tax burden designed for OECD countries and constructed as a combination of hard tax data weighted by soft data (QEO) gathered through an online questionnaire survey among tax specialists. In 2016 we introduced new methodology of WTI based on factor analysis and pilot project was held in 2017 for the Czech Republic to test the relevance of the method. In this paper we present and discuss the results of pilot project. We use three different factorization methods (namely Principal Component Analysis, Factor Analysis and Non-negative Matrix Factorization) to find latent factors in data and afterward compute new WTI values for each factorization method. All three methods suggest that corporate and personal income taxes have the highest tax burden effect.

Keywords: Tax Burden, World Tax Index (WTI), Tax Quota, Principal Component Analysis, Factor Analysis, Non-negative Matrix Factorization, Czech Republic.

JEL classification: H21, C38
AMS classification: 62H25, 91B64

1 Introduction

Fiscal policy is built on two main instruments that complement each other. Every government needs to cover its spending by collecting taxes. Taxes may have negative impact on economy and slow down economic growth as they create distortions and burden part of economic activity. It is crucial task for politicians to find equilibrium in which they collect sufficient funds to cover their activities without burdening private sector too much. This task is even more important nowadays. Governments are forced to consolidate public finances as they are still struggling with the consequences of the latest global financial crisis.

Taxes are the essential parts of models of economic growth, therefore we need to be able measure the size of tax burden correspondingly. The goal of this paper is to present the results of the pilot project in the Czech Republic. Firstly, we will briefly present the different measurements of tax burden, then we will explain the details of calculating the World Tax Index based on new methodology. We will compare the results obtained by using three different approaches and discuss the relevance of the new methodology.

2 Tax Burden Measurements

Tax quota and implicit tax rates are the mostly used indices to measure tax burden. The first mentioned, tax quota simply expresses the ratio of tax revenues to nominal GDP, or in compound form ratio of the tax revenues with social security contributions to nominal GDP. The indicator is often used for comparison and in models of economic growth because is simple and available for many countries. However, according to Kotlán & Machová [8], the tax quota reflects only the percentage of GDP redistributed through the public budgets rather than the level of the tax burden. A higher tax quota can indicate only more efficient tax collection and not the higher tax burden. Another disadvantage is a usage of GDP that has different methodology in countries.

Implicit tax rate is the second popular indicator of tax burden. It again relates the tax revenues but only to the activities or commodities in GDP connected to the tax instead of the whole nominal GDP. However, this change does not eliminate the disadvantages mentioned above. The implicit tax rates are available for consumption, labour and capital but only for European Union and therefore international comparison is not possible outside European Union. More about disadvantages of using these two indicators can be found in Kotlán et al. [7].

We can find also alternative indicators of tax burden. Raimondos-Møller & Woodland [12] present a non-money metric index of optimality that measures the distance of some current tax structure from the optimal tax structure

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno, konopkova.zlatica@mail.muni.cz
[2]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno, jakubbucek@mail.muni.cz

in the presence of public goods. De Laet & Wöhlbier [2] define a tax as a compulsory unrequired payment to the government, differentiate the types of taxes according to the national account and assigned them various economic functions. Kiss et al. [5] eliminates the influence of government in tax quota.

World Tax Index (WTI) is an overall multi-criteria indicator of the tax burden designed for OECD countries. It was firstly presented in Kotlán & Machová [8] as an alternative for standardly used indicators (tax quota, implicit tax rates). WTI is not limited only to the tax revenues and nominal GDP, but it includes information from different areas that can possibly affect the tax burden, e.g. progressivity of the tax system, administration, incentives, deductions etc. The results presented in Kotlán & Machová [9] show that the WTI generally allows to prove more significantly and clearly the negative effect of taxation in case of particular types of taxes, which is not possible using the tax quota.

WTI is structured according to OECD classification into 5 main groups of taxes - Personal Income Tax (PIT), Corporate Income Tax (CIT), Value Added Tax (VAT), Individual Property Taxes (PRO) and Other Taxes on Consumption/Excises (OTC). The values of WTI are computed as a combination of available hard tax data with soft data (Qualified Expert Opinion QEO) gained from questionnaire survey conducted in all OECD countries. In the initial verstion of WTI, the questionnaire was distributed among tax specialists in academic sphere. They expressed their opinion on how individual components of WTI contribute to the tax burden in their country by distributing the 100% in every major category. The Saaty method[3] of pair-wise comparisons was then applied to obtain the weights of categories. In Konôpková & Buček [6] we introduced new methodology in order to diminish potential disadvantages.

3 Methodology and data

First of all, the initial method of WTI was built on assumption that experts can evaluate the tax burden in their country directly. According to Borsboom et al. [1] human behavior cannot be observed directly and the variables expressing the behavior are latent, their values are hidden for us, but we can use a mathematical method to estimate them. Secondly, the questionnaire was distributed only to experts in academic sphere, but the weights should reflect proportionally the opinion of the whole diverse population in country. We have chosen factor analysis to obtain the weights of tax categories in WTI and have opened the questionnaire to public and private sector. In this section we will present three different methods of factorization that was used to check the robustness of our results (to evaluate the relevance of the new methodology) and the data.

3.1 Methodology

Let X be a $n \times p$ matrix. In our case, n is number of respondents (rows of X) and p is number of questions (columns of X). We want to find the latent effect using factorization of the matrix X. The main principle of the factorization is to decompose a matrix into two parts: *a basis matrix* with new basis and *a transformation matrix* that transforms data from original basis to the new ones. Since the proportion of variance for each new base to total data's variance is known, one can reduce dimension of data by removing basis with the lowest variance.

Applying factorization procedure to data we will find 5 factors, one factor for each tax category. Tax categories will be assigned to every factors on basis of correlation between factors and corresponding questions from questionnaire. Afterwards new scores for every observations will be computed. Scores for each observation will be transformed to scale 0-1 using following formula:

$$s_{ij}^T = \frac{s_{ij}^O - \min\{s_{i1}^O, s_{i2}^O, s_{i3}^O, s_{i4}^O, s_{i5}^O\}}{\max\{s_{i1}^O, s_{i2}^O, s_{i3}^O, s_{i4}^O, s_{i5}^O\} - \min\{s_{i1}^O, s_{i2}^O, s_{i3}^O, s_{i4}^O, s_{i5}^O\}}, \tag{1}$$

where s_{ij}^O denotes the j-th original score for the i-th respondent and s_{ij}^T is the j-th transformed score for the i-th respondent. So the tax category that causes the highest tax burden to the i-th individual will be 1 and the tax category with lowest tax burden will be 0.

After that, individual weights will be computed as follows. Firstly, an average score for each individual will be computed as $s_i = \frac{1}{5}(s_{i1}^O + s_{i2}^O + s_{i3}^O + s_{i4}^O + s_{i5}^O)$. Then the weight for the i-th individual will be calculated using following formula:

$$w_i = \left| \frac{s_i - mean(s_i)}{\sqrt{var(s_i)}} \right|^{-1}. \tag{2}$$

Finally, weights for each tax category will be computed as weighted mean of all individuals transformed scores,

[3] See e.g. Saaty [13].

i.e. the weight for the j-th WTI category is $\sum_i w_i \cdot s_{ij}^T$. All weights of WTI categories are normalized to give sum of 1. This process will be repeated for every method mentioned below.

Principal Component Analysis

The Principal Component Analysis (PCA) is a statistical procedure that transforms a number of correlated variables into a (smaller) number of uncollerated variables. The PCA can be done using the singular value decomposition (SVD):

$$X = U\Sigma V^T, \tag{3}$$

where U is a $n \times n$ matrix with eigenvectors of XX^T, Σ is a $n \times p$ diagonal matrix containing the square roots of eigenvalues of U or V in descending order and V^T is a $p \times p$ matrix with eigenvectors of X^TX.

Factor analysis

Factor Analysis (FA) is similar to PCA, but different from it, the FA seeks for a latent structure in the data. The FA assumes that variables $X_1, X_2, \ldots, X_p$ (columns of matrix X) are linear function of m ($m < p$) underlying factors $F_1, F_2, \ldots, F_m$. In matrix notation, factor analysis can be described by the following equation

$$C = PQP^T + U^2, \tag{4}$$

where C is a $p \times p$ correlation matrix among observed variables, P is a $p \times m$ loading matrix, Q is a $m \times m$ matrix of correlations among factors and U^2 is a $p \times p$ diagonal matrix of unique variances. Factors can be rotated for better interpretation.

Non-negative matrix factorization

Assume that the matrix X is non-negative, i.e. $x_{ij} \geq 0$. Non-negative Matrix Factorization (NMF) consists in finding an approximation

$$X \approx WH, \tag{5}$$

where W is a $n \times r$ non-negative matrix called *basis matrix* and H is a $r \times p$ non-negative matrix called *mixture coefficient matrix*. The non-negativity property makes matrix easier to interpret. The NMF in not unique as one can find a matrix B and its inverse B^{-1} such as $\tilde{W} = WB$ and $\tilde{H} = B^{-1}H$. If the new matrices are non-negative they from another parametration of the factorization. The NMF was calculated using R package by Gaujoux et al. [4].

3.2 Data

Data were collected in period from October 2016 till April 2017. Questionnaire was sent to academic researchers, public national institutions and private sector represented by the Chamber of Tax Consultants of the Czech Republic and the 'Big Four' group. They were asked to answer "Do you think...?" questions in standard Likert scale *Strongly disagree* (1) - *Strongly agree* (5). Questions and their median answers are displayed in Table 3.

To calculate WTI for the Czech Republic we have chosen the representing tax rate/rates (see Table 1). Their values and changes over time are shown in the Figure 1. The most significant changes are in OTC and CIT caused by the entry of the Czech Republic into the European Union in 2004 when the tax harmonization has started.

Category	Description	Source
PIT	Average personal income tax & soc. sec. contribution rates on gross labour income (100% average wage)	OECD [11] Table I.5
CIT	Statutory corporate income tax rate	OECD [11] Table II.1
VAT	Basic tax rate	OECD [11] Table 2.A2.1
PRO	Tax on the acquisition of immovable property/property transfer tax + land tax	Finance.cz [3]
OTC	Excises average rate weighted by consumption	OECD [11] Table I.5

Table 1 Representing tax rates for the Czech Republic and their source.

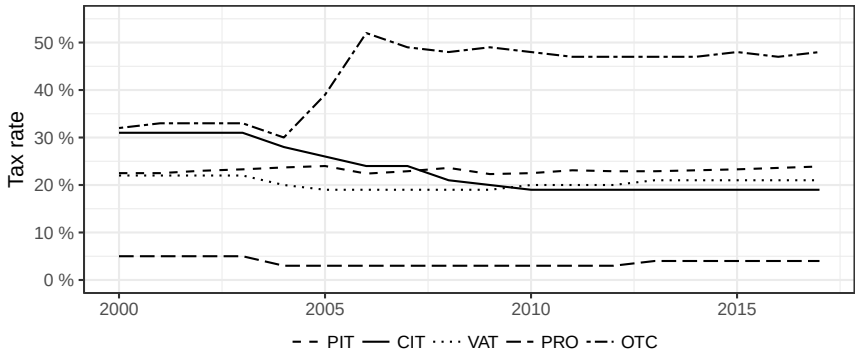


Figure 1 Representing tax rates in time.

4 Results

Together 54 respondents filled the questionnaire, most of them from the private sphere (80%). The public sphere did not respond too much (only 3 responses) and did not even respond to our repeating emails. Almost 75% respondents strongly agree or agree with the statement that they consider themselves as experts on taxes, less than 10% strongly disagree or disagree with it and remaining 15% are neutral.

Applying methodology mentioned above we received the weights and order of the categories in WTI displayed in Table 2. As we can see the most important in WTI is Corporate Income Tax along with Personal Income Tax. Both categories fall within direct taxes which can be the reason why people find it the most influencing. The order of other categories varies depending on used method. By two of three methods, the third highest weight belongs to Other Taxes on Consumption/Excises. Excises are the highest taxes in the Czech Republic and are imposed on desired goods The last categories are Value Added Tax and Property Taxes. The results are in line with OECD [10] which supports the main trend in fiscal policy, to make a move from direct taxes to indirect ones. It states that corporate and personal income taxes have the largest negative effects on economy and the consumption and property taxes are on the opposite side.

	Weights			**Order**		
Category	**PCA**	**FA**	**NMF**	**PCA**	**FA**	**NMF**
PIT	0.209	0.224	0.246	2	2	2
CIT	0.302	0.242	0.247	1	1	1
VAT	0.170	0.163	0.140	4	4	5
PRO	0.102	0.159	0.190	5	5	3
OTC	0.197	0.212	0.177	3	3	4

Table 2 Calculated weights and order of the categories in WTI.

Finally, we calculated the values of new WTI for all three methods, WTI_{FA}, WTI_{PCA},WTI_{NMF}. They are displayed in Figure 2 together with Tax Quota and old values of WTI_{OLD}[4]. The indicator of tax burden defined in this way is actually a single weighted average tax rate for the country without the influence of GDP. As we can see, the values of new WTI are independent of chosen factorization method. In comparison with WTI_{OLD} it better reflects the real state of tax system. It does not overestimate tax burden and there are no significant fluctuations that have no support in real data.

5 Conclusion

In this paper we compared three different factorization methods (PCA, FA and NMF) for computing new values of World Tax Index for the Czech Republic for period 2000–2017. The indicator of tax burden defined in this way is actually a single weighted average tax rate for the country. All methods get similar effects of the five tax categories and therefore the computation of the WTI is independent of chosen method of factorization. According to our results, Corporate and Personal Income Tax are the most important categories of WTI, following by Other Taxes on Consumption/Excises. The least important categories are Value Added Tax and Property Taxes. These findings are in line with literature. The new WTI is more stable, without any significant breaks compared to the old WTI. It also has noticeable lower values and more corresponds to the tax quota which is standard tax burden indicator.

[4]Available at https://data.oecd.org/tax/tax-revenue.htm and http://www.eaco.eu/about-eaco/research/.

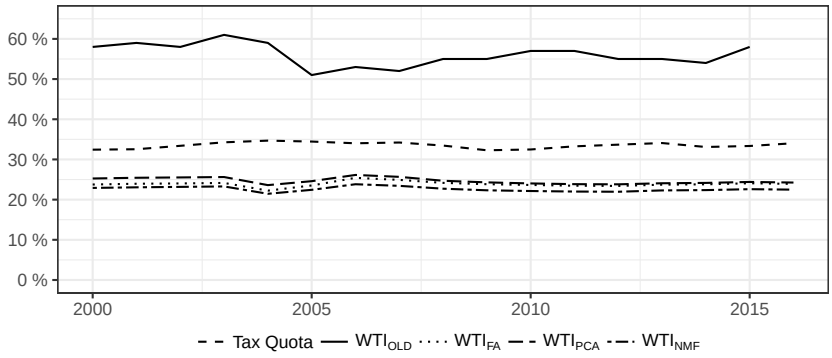


Figure 2 Tax burden measurements.

Acknowledgements

This work was supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/1139/2017. This support is gratefully acknowledged.

References

[1] Borsboom, D., Mellenbergh, G. J. & van Heerden, J. (2003). The Theoretical Status of Latent Variables. *Psychological Review*, 110(2), pp. 203-219.

[2] De Laet, J.-P. & Wöhlbier, F. (2008). Tax Burden by Economic Function: A Comparison for the EU Member States. *MPRA Paper*, 14761.

[3] [6] Finance.cz, (2017) [online]. Available at: https://www.finance.cz/dane-a-mzda/majetkove-dane/dan-z-prevodu-nemovitosti/ [Accessed 5 Oct. 2017].

[4] Gaujoux, R. & Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bionformatics 2010*, 11:367.

[5] Kiss, G. P., Jedrzejowicz, T. & Jirsáková, J. (2009). How to Measure Tax Burden in an Internationally Comparable Way? *National Bank of Poland Working Paper*, 56.

[6] Konôpková, Z. & Buček, J. (2016). World Tax Index: Methodology for 2017. In Machová: *6th International Scientific Conference TAXES IN THE WORLD*. Krakow: Ostrava: European Association Comenius EACO and VSEO, pp. 77-87.

[7] Kotlán, I., Machová, Z. & Janíčková, L. (2011). Taxation Influence on the Economic Growth. *Politická ekonomie*, 59(5), pp. 638658.

[8] Kotlán, I. & Machová, Z. (2012). World Tax Index: Methodology and Data. *DANUBE: Law and Economics Review*, 3(2), pp. 1933.

[9] Kotlán, I. & Machová, Z. (2013). The Impact of the Tax Burden on the Living Standard in OECD Countries. *Ekonomický časopis*, 61(9), pp. 951962.

[10] OECD, (2011). *Tax Policy Reform and Economic Growth*. Washington: OECD.

[11] OECD, (2017). Database OECD. [online]. Available at: http://www.oecd.org/tax/tax-policy/tax-database.htm [Accessed 5 Oct. 2017].

[12] Raimondos-Møller, P. & Woodland, A. D., (2006). Measuring Tax Efficiency: A tax optimality index. *Journal of Public Economics*, 90(12), pp. 1903-1922.

[13] Saaty, T. L. (2008). Decision Making with the Analytic Hierarchy Process. *International Journal of Services Sciences*, 1(1), pp. 8398.

Appendix A

ID	Do you think that...	Median answer
Personal Income Tax		
B1	the tax literacy is accurate in your country?	Disagree
B2	the progressivity of tax system affect the individuals decision?	Disagree
B3	the tax personal deductions are appropriately created and used by people?	Disagree
B4	the social security contributions help to reduce the inequality in your country?	Disagree
B5	lot of people in your country avoid paying taxes by shift to shadow economy?	Disagree
Corporate Income Tax		
A1	lot of firms in your country have location in tax havens?	Disagree
A2	the progressivity of tax system affects the firms decision?	Disagree
A3	the tax system influences the firms market entry?	Strongly disagree
A4	the tax deductibility of costs plays important role in firms finances?	Strongly disagree
A5	the corporate income taxes influence the firm's investment?	Strongly disagree
Value Added Tax		
C1	the value added tax influence the consumption of durable goods?	Disagree
C2	the reduced tax rates are useful for the most of population?	Disagree
C3	people in your country consider the double prices (VAT included/excluded) on price tags?	Neutral
C4	VAT motivates people to buy products outside the official market?	Disagree
C5	it is complicated for firms to fulfill the mandatory registration duty for VAT?	Disagree
Property Taxes		
D1	the taxes on immovable property influence the type of accommodation?	Neutral
D2	the taxes on net wealth influence the place of residence?	Neutral
D3	the inheritance and gift taxes motivate people to spend their wealth on their own benefit instead of saving for their children?	Disagree/Neutral
D4	the people in your country think more about the future than present?	Disagree
D5	the taxes on financial and capital transactions discourage people from buying bonds and shares?	Disagree
Excises		
E1	the increase in excise influence the consumption of alcohol?	Disagree
E2	the increase in excise influence the consumption of tobacco?	Disagree
E3	the increase in excise influence the consumption of mineral oils?	Disagree
E4	your country is "green" country?	Disagree
E5	there is problem with civilization diseases in your country?	Neutral

Table 3 Questions

Endogenous randomness and first–order stochastic dominance in portfolio optimization.

Miloš Kopa[1]

Abstract. The paper deals with portfolio selection problems which maximize mean portfolio return under constraints that the random return outperform a random benchmark. The outperformance can be understood in several different ways. In this paper, we focus on the first-order stochastic dominance constraints. Firstly, we modify the classical first-order stochastic dominance relation between returns of two given portfolios for the case with endogenous randomness of returns. Endogenous randomness (or decision dependent randomness) means that the probability distribution of asset returns may depend on the decision variable, i.e. on the weights associated to the assets. This is applicable mainly in the high frequency trading, emission allowance trading and/or in the illiquid markets. Secondly, we apply this new portfolio selection rule in the portfolio optimization problem.

Keywords: Endogenous randomness, first–order stochastic dominance, portfolio optimization

JEL classification: D81, G11
AMS classification: 91B16, 91B30

1 Introduction

Uncertainty in stochastic optimization models is distinguished in two cases: endogenous and exogenous. Endogenous uncertainty is inner uncertainty of the model. It means that the random (uncertain) element of the problem may depend on the solution (decision). For example, the decision-maker can force one possibility to become more probable. Exogenous uncertainty is entering the model from outside. The decision-maker cannot affect the underlying probability distribution. The previous work on the endogenous uncertainty in optimization problems is limited to a few papers only. Their reviews can be found e.g. in [26], [5] or [9].

The most commonly used stochastic dominance relations are so called the first-order and the second-order stochastic dominance. They allow to compare random variables, in financial applications random returns or losses. We say that a random variable X dominates a random variable Y with respect to the N-th order stochastic dominance, $N = 1, 2$ if

$$\mathbb{E}u(X) \geq \mathbb{E}u(Y) \quad \text{for all} \quad u \in U_N$$

where the set of admissible utility functions for the N-th order stochastic dominance is defined as follows:

$$U_N = \{u(x) \in D_N : (-1)^k u^{(k)} \leq 0, \quad \forall x, \quad k = 1, ..., N\}$$

where D_N is the class of N-times differentiable functions and $u^{(k)}$ stands for the k-th derivative of function u.

In particular, the first-order stochastic dominance (FSD) is generated by the set of all non-decreasing functions (all non-satiated decision makers are considered), while the second order stochastic dominance (SSD) focuses only on the risk averters, that is, only concave and non-decreasing functions are admissible. The basics go back to 1960th, see [23], [10], [11] or [25]. In these papers, several sufficient and necessary conditions for the N-th order stochastic dominance, $N = 1, 2$, were derived. In particular a random variable X dominates a random variable Y with respect to FSD if and only if

$$F_X(z) \leq F_Y(z) \quad \forall z \in \mathbb{R}$$

[1]The Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodrenskou v 4, CZ-182 08, Prague 8, Czech Republic, kopa@karlin.mff.cuni.cz

where $F_X(z)$ is the cumulative distribution function of X. Similarly, a random variable X dominates a random variable Y with respect to the second-order stochastic dominance if and only if

$$F_X^{[2]}(z) \leq F_Y^{[2]}(z) \quad \forall z \in \mathbb{R}$$

where

$$F_X^{[2]}(z) = \int_{-\infty}^{z} F_X(t)dt.$$

See [16] for more details. In last years, a substantial development of the stochastic dominance applications in finance was observed mainly for the second–order stochastic dominance which is computationally more tractable. In particular:

- necessary and sufficient conditions for portfolio efficiency with respect to stochastic dominance criteria were derived and applied, see e.g. [21], [8], [15], [20]
- portfolio enhancement using stochastic dominance rules proved to be a promissing way comparing to classical mean-risk models, see e.g. [24], [12], [22], [14]
- Special Data Envelopment Analysis models wich are equivalent to portfolio efficiency tests with respect to stochastic dominance were presented in [1], [2], [3].
- more robust version of stochastic dominance relations and stochastic dominance efficiency were proposed, see e.g. [4], [13], [6] or [7].
- more general stochastic dominance (ordering) rules were introduced and analyzed, see e.g. [18], [17] and references therein, or [19].

Since the attention was paid mainly to the second-order stochastic dominance, in this paper we focus on the portfolio selection problems which maximize expected return under the first-order stochastic dominance constraints. First, we formulate a portfolio enhancement model with respect to the first-order stochastic dominance under assumption of exogenous randomness. Then we modify the model for the case of decision dependent randomness. For both cases we present some tractable formulations under assumption that the distributions of returns are discrete. Especially we focus on the case where the realizations of the random variables are equiprobable in order to present the ideas in the simplest way. Finally, we present a numerical example showing how the optimal portfolios may differ if endogenous randomness is incorporated in the problem.

The remainder of this paper is structured as follows. Section 2 presents a notation and basic properties of the FSD relation. It is followed by a summary of portfolio selection models (with exogenous or endogenous randomness) applying FSD constraints in Section 3. These models assume an empirical distribution of the returns. Section 4 shows a numerical example. The paper is concluded in Section 5.

2 Preliminaries and notations

We consider a random vector $\mathbf{r} = (r_1, r_2, ..., r_N)'$ of returns of N assets in T scenarios taken with probabilities p_t. The returns of the assets for the various scenarios are given by

$$X = \begin{pmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^T \end{pmatrix}$$

where $\mathbf{x}^t = (x_1^t, x_2^t, \ldots, x_N^t)$ is the t-th row of matrix X. We will use $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)'$ for a vector of portfolio weights and the portfolio possibilities are given by

$$\Lambda = \{\boldsymbol{\lambda} \in R^N | \mathbf{1}'\boldsymbol{\lambda} = 1, \quad \lambda_n \geq 0, \quad n = 1, 2, \ldots, N\}.$$

Let $F_{\mathbf{r}'\boldsymbol{\lambda}}(x)$ denote the cumulative probability distribution function of returns of portfolio $\boldsymbol{\lambda}$.

Definition 1. Portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates portfolio $\boldsymbol{\tau} \in \Lambda$ by the first-order stochastic dominance ($\mathbf{r}'\boldsymbol{\lambda} \succ_{FSD} \mathbf{r}'\boldsymbol{\tau}$) if

$$F_{\mathbf{r}'\boldsymbol{\lambda}}(t) \leq F_{\mathbf{r}'\boldsymbol{\tau}}(t) \qquad \forall t \in \mathbb{R}.$$

This relation is sometimes called a weak FSD and the equivalent definition, presented in e.g. [11], [16] or [15] is based on comparison of expected utility of portfolio returns:

$$\mathbf{r}'\boldsymbol{\lambda} \succ_{FSD} \mathbf{r}'\boldsymbol{\tau} \Longleftrightarrow \mathbb{E}u(\mathbf{r}'\boldsymbol{\lambda}) \geq \mathbb{E}u(\mathbf{r}'\boldsymbol{\tau})$$

for all utility functions $u \in U_1$.

Following [7] we may formulate the FSD constraints using Value at Risks as follows. Assume that $p_t^{\boldsymbol{\lambda}}$ are probabilities ordered according to the corresponding values of returns of portfolio $\boldsymbol{\lambda}$. Let $q_s^{\boldsymbol{\lambda}}$ are cumulative probabilities for portfolio $\boldsymbol{\lambda}$, i.e.

$$q_s^{\boldsymbol{\lambda}} = \sum_{t=1}^{s} p_t^{\boldsymbol{\lambda}}$$

and the same notation is applied to the portfolio $\boldsymbol{\tau}$. Then

$$\mathbf{r}'\boldsymbol{\lambda} \succ_{FSD} \mathbf{r}'\boldsymbol{\tau} \Longleftrightarrow \mathrm{VaR}_{\mathrm{q}_\mathrm{s}^{\boldsymbol{\tau}}}(-\mathbf{r}'\boldsymbol{\lambda}) \geq \mathrm{VaR}_{\mathrm{q}_\mathrm{s}^{\boldsymbol{\tau}}}(-\mathbf{r}'\boldsymbol{\tau})$$

where $\mathrm{VaR}_{\mathrm{q}_\mathrm{s}^{\boldsymbol{\tau}}}(-\mathbf{r}'\boldsymbol{\lambda}) = \min\{\delta : \mathrm{P}(-\mathbf{r}'\boldsymbol{\lambda} \leq \delta) \geq \mathrm{q}_\mathrm{s}^{\boldsymbol{\tau}}\}$.

In a special case, when $p_t = 1/T$ the necessary and sufficient condition for FSD can be rewritten using a permulation matrix P as follows:

$$\mathbf{r}'\boldsymbol{\lambda} \succ_{FSD} \mathbf{r}'\boldsymbol{\tau} \tag{1}$$

$$\Longleftrightarrow$$

$$\exists P = \{p\}_{i,j=1}^{T} : X\boldsymbol{\lambda} \geq PX\boldsymbol{\tau}, \sum_{i=1}^{T} p_{i,j} = 1, j = 1, 2, ..., T, \sum_{j=1}^{T} p_{i,j} = 1, i = 1, 2, ..., T, p_{i,j} \in \{0, 1\}. \tag{2}$$

If we allow that $\mathbf{r}$ may depend on the decision vector $\boldsymbol{\lambda}$ we have to make Definition 1 more general.

Definition 2. Portfolio $\boldsymbol{\lambda} \in \Lambda$ dominates portfolio $\boldsymbol{\tau} \in \Lambda$ by the first-order stochastic dominance with endogenous randomness ($\mathbf{r}(\boldsymbol{\lambda})'\boldsymbol{\lambda} \succ_{FSD} \mathbf{r}(\boldsymbol{\tau})'\boldsymbol{\tau}$) if

$$F_{\mathbf{r}(\boldsymbol{\lambda})'\boldsymbol{\lambda}}(t) \leq F_{\mathbf{r}(\boldsymbol{\tau})'\boldsymbol{\tau}}(t) \qquad \forall t \in \mathbb{R}.$$

If endogenous randomness affects only the values of the return scenario matrix X but not the probabilities of the scenarios we may modify (1)-(2) as follows:

$$\mathbf{r}(\boldsymbol{\lambda})'\boldsymbol{\lambda} \succ_{FSD} \mathbf{r}(\boldsymbol{\tau})'\boldsymbol{\tau} \tag{3}$$

$$\Longleftrightarrow$$

$$\exists P = \{p\}_{i,j=1}^{T} : X(\boldsymbol{\lambda})\boldsymbol{\lambda} \geq PX(\boldsymbol{\tau})\boldsymbol{\tau}, \sum_{i=1}^{T} p_{i,j} = 1, j = 1, 2, ..., T, \sum_{j=1}^{T} p_{i,j} = 1, i = 1, 2, ..., T, p_{i,j} \in \{0, 1\}. \tag{4}$$

where $X(\boldsymbol{\lambda})$ is the scenario return matrix if portfolio $\boldsymbol{\lambda}$ is chosen. Similarly, if we invest in portfolio $\boldsymbol{\tau}$ the scenario return matrix is $X(\boldsymbol{\tau})$.

3 Portfolio optimization models

In this paper, we are dealing with portfolio selection models with stochastic dominance constraints. If exogenous randomness is assumed, the model takes the form:

$$\begin{aligned} & \max \mathbb{E}(\mathbf{r}'\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \mathbf{r}'\boldsymbol{\lambda} \succ_{FSD} \mathbf{r}'\boldsymbol{\tau} \\ & \boldsymbol{\lambda} \in \Lambda \end{aligned} \tag{5}$$

where $\boldsymbol{\tau}$ is a given benchmark (reference) portfolio. If we consider a discrete distribution of returns with equiprobable scenarios (atoms) we can follow (1)-(2) and formulate it as a linear mixed integer programming problem:

$$\max \mathbb{E}(\mathbf{r}'\boldsymbol{\lambda}) \tag{6}$$

$$\text{s.t.} \quad X\boldsymbol{\lambda} \geq PX\boldsymbol{\tau}, \tag{7}$$

$$\sum_{i=1}^{T} p_{i,j} = 1, j = 1, 2, ..., T, \tag{8}$$

$$\sum_{j=1}^{T} p_{i,j} = 1, i = 1, 2, ..., T, \tag{9}$$

$$p_{i,j} \in \{0, 1\} \tag{10}$$

$$\boldsymbol{\lambda} \in \Lambda \tag{11}$$

In the case, that endogenous randomness affects only the values of the return scenario matrix X but not the probabilities of the scenarios, following (3)-(4) we get:

$$\max \mathbb{E}(\mathbf{r}(\boldsymbol{\lambda})'\boldsymbol{\lambda}) \tag{12}$$

$$\text{s.t.} \quad X(\boldsymbol{\lambda})\boldsymbol{\lambda} \geq PX(\boldsymbol{\tau})\boldsymbol{\tau}, \tag{13}$$

$$\sum_{i=1}^{T} p_{i,j} = 1, j = 1, 2, ..., T, \tag{14}$$

$$\sum_{j=1}^{T} p_{i,j} = 1, i = 1, 2, ..., T, \tag{15}$$

$$p_{i,j} \in \{0, 1\} \tag{16}$$

$$\boldsymbol{\lambda} \in \Lambda \tag{17}$$

4 Numerical example

In this section we assume only three assets with three equiprobable scenarios of returns (in percentage). In particular

$$X = \begin{pmatrix} 3 & 8 & 5 \\ 2 & 4 & 3 \\ 11 & 3 & 6 \end{pmatrix}$$

and the benchmark portfolio is $\boldsymbol{\tau} = (0, 0, 1)$. Let's solve first the problem (8)-(13), that is, the model with exogenous randomness. In this case, it is easy to see that the first asset is the most profitable, so we would like to invest in it as much as possible. Since the minimal return of the benchmark portfolio is 3% the maximal possible (and optimal) weight of the first asset is 0.5 when combined with the second asset. Hence, the optimal portfolio $\boldsymbol{\lambda}^* = (0.5, 0.5, 0)$ and the optimal permutation matrix is the identity matrix.

Now assume that if a massive investment is done in the first (the second) asset, that is $\lambda_1 \geq 0.8$ ($\lambda_2 \geq 0.8$) the returns of the asset are increased by 1 in all scenarios. Consider the following subsets of Λ:

$$\Lambda_1 = \{\boldsymbol{\lambda} \in \Lambda : \lambda_1 \geq 0.8\}$$

$$\Lambda_2 = \{\boldsymbol{\lambda} \in \Lambda : \lambda_2 \geq 0.8\}.$$

Now we can express the decision dependent random returns:

$$X(\boldsymbol{\lambda}) = \begin{pmatrix} 3 & 8 & 5 \\ 2 & 4 & 3 \\ 11 & 3 & 6 \end{pmatrix} + \mathcal{I}(\boldsymbol{\lambda} \in \Lambda_1) \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + \mathcal{I}(\boldsymbol{\lambda} \in \Lambda_2) \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

In this case, problem (14)-(19) can be decomposed in three separate problems which are very similar to each other, just the constraint (19) is modified as follows:

1. $\boldsymbol{\lambda} \in \Lambda_1$
2. $\boldsymbol{\lambda} \in \Lambda_2$
3. $\boldsymbol{\lambda} \in \Lambda \backslash (\Lambda_1 \cup \Lambda_2)$.

When (19) is replaced by $\boldsymbol{\lambda} \in \Lambda_1$ no feasible portfolio exists because the second largest return of any $\boldsymbol{\lambda} \in \Lambda_1$ is always smaller than the second largest return of the benchmark portfolio. In the second case, the optimal portfolio is $\boldsymbol{\lambda} = (0, 1, 0)$ and the optimal objective value is 6. The third modification when $\boldsymbol{\lambda} \in \Lambda \backslash (\Lambda_1 \cup \Lambda_2)$ gives the same optimal solution as the exogenous randomnes case $\boldsymbol{\lambda} = (0.5, 0.5, 0)$ and its optimal objective value 31/6. Since $6 > 31/6$ the optimal solution of the endogenous randomness portfolio selection model is $\boldsymbol{\lambda}^* = (0, 1, 0)$.

5 Conclusions

In this paper we analyzed the portfolio selection models with the first-order stochastic dominance constraints under assumption of the endogenous randomness. It means that the multivariate distribution of random returns depends on the decision vector (portfolio weights). We first introduce a definition of FSD between two portfolios under assumption of endogenous randomness and then we formulate the portfolio selection model assuming that only the values of scenarios may depend on the decision vector but not the probabilities of the scenarios. Moreover, we focus on the simpliest case when all scenarios are equiprobable. Finally, we illustrated a difference between endogenous and exogenous randomness portfolio selection problem on a toy example considering only three scenarios and three assets. Despite the fact that the endogenous randomness is expressed in a simple way, the optimal porfolio of the endogenous randomness problem very much differ from the one of the exogenous randomness problem. Therefore, omitting the endogenous feature of the returns would lead to completely different conclusions.

For the future research, this study can be improved in various ways. For example, one can consider the stochastic dominance relations in a more robust way as it was done in [13] or [6], [7], using contamination techniques and the worst-case approach. Alternatively, one can compare the results also with the case when the second-order or the third-order stochastic dominance is used.

Acknowledgements

The paper was supported by the grant No. 16-01298S of the Czech Science Foundation.

References

[1] Branda, M., and Kopa, M.: DEA-risk Efficiency and Stochastic Dominance Efficiency of Stock Indices. *Czech Journal of Economics and Finance* **62** (2012), 106–124.

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

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

[4] Dentcheva, D., and Ruszczyński, A.: Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming, Series B* **123** (2010), 85–100.

[5] Dupačová, J.: Optimization under exogenous and endogenous uncertainty. In: Proc. of MME06 (L. Luk, ed.), University of West Bohemia 2006, 131-136.

[6] Dupačová, J., and Kopa, M.: Robustness in stochastic programs with risk constraints. *Annals of Operations Research* **200** (2012), 55–74.

[7] Dupačová, J., and Kopa, M.: Robustness of optimal portfolios under risk and stochastic dominance constraints. *European Journal of Operational Research* **234** (2014), 434–441.

[8] Grechuk, B.: A simple SSD-efficiency test. *Optimization Letters* **8** (2014), 2135–2143.

[9] Grossmann, I.E, Apap, R.M., Calfa, B.A., Garca-Herreros, P. and Zhang, Q.: Recent advances in mathematical programming techniquesfor the optimization of process systems under uncertainty. *Computers and Chemical Engineering* **91** (2016), 3-14.

[10] Hadar, J., and Russell, W. R.: Rules for Ordering Uncertain Prospects. *American Economic Review* **59** (1969), 25–34.

[11] Hanoch, G., and Levy, H.: The efficient analysis of choices involving risk. *Review of Economic Studies* **36** (1969), 335–346.

[12] Hodder, E., Jackwerth, J.C., and Kolokolova, O.: Improved Portfolio Choice Using Second-Order Stochastic Dominance. *Review of Finance* **19** (2015), 1623–1647.

[13] Kopa, M.: Measuring of second-order stochastic dominance portfolio efficiency. *Kybernetika* **46** (2010), 488–500.

[14] Kopa, M., Moriggia, V., Vitali, S.: Individual optimal pension allocation under stochastic dominance constraints. *Annals of Operations Research* **260** (2018), 255–291.

[15] Kopa, M., and Post, T.: A general test for SSD portfolio efficiency. *OR Spectrum* **37** (2015), 703–734.

[16] Levy, H.: *Stochastic dominance: Investment decision making under uncertainty.* Second edition, Springer, New York, 2006.

[17] Ortobelli, S., Lando, T., Petronio, F., and Tichy, T.: Asymptotic stochastic dominance rules for sums of i.i.d. random variables. *Journal of computational and applied mathematics* **300** (2016), 432–448.

[18] Ortobelli, S., Shalit, H., and Fabozzi, F.: Portfolio selection problems consistent with given preference orderings. *International Journal of Theoretical and Applied Finance* **16** (2013), 38p.

[19] Post, T.: Standard Stochastic Dominance. *European Journal of Operational Research* **248** (2016), 1009–1020.

[20] Post, T., Fang, Y., and Kopa, M.: Linear Tests for DARA Stochastic Dominance. *Management Science* **61** (2015), 1615–1629.

[21] Post, T., and Kopa, M.: General Linear Formulations of Stochastic Dominance Criteria. *European Journal of Operational Research* **230** (2013), 321–332.

[22] Post, T., and Kopa, M.: Portfolio Choice based on Third-degree Stochastic Dominance. *Management Science* **63** (2017), 3381–3392 .

[23] Quirk, J. P., and Saposnik, R.: Admissibility and measurable utility functions. *Review of Economic Studies* **29** (1962), 140-146.

[24] Roman, D., Mitra, G., and Zverovich, V.: Enhanced indexation based on second-order stochastic dominance. *European Journal of Operational Research* **228** (2013), 273–281.

[25] Rothschild, M., and Stiglitz, J.E.: Increasing risk: I. A definition. *Journal of Economic Theory* **2** (1970), 225–243.

[26] Sahinidis, N. V.: Optimization under uncertainty: State-of-the-art and opportunities. *Computers and Chemical Engineering* **28** (2004), 971-983.

Application of the *p*-Median Approach for a Basic Decomposition of a Set of Vertices to Service Vehicles Routing Design

Petr Kozel[1], Lucie Orlíková[2], Marek Pomp[3], Šárka Michalcová[4]

Abstract. In order to design the service vehicle routes the main goal is to minimize the total distance travelled (in kilometers). In addition to this criterion, it is necessary to take into account a lot of other constraints, which are common in a real traffic. For example, one of the basic requirements is not to exceed the service vehicle capacity. Moreover, another requirement may be creating a defined number of subsets, in which service requirements can consequently be realized. The presented text deals with the application of the *p*-median approach for a basic decomposition of a set of vertices to partial subsets, where service vehicle routes are identified by using of a suitable algorithm (an exact algorithm for small tasks, a heuristic algorithm for large tasks). The computational experiments presented in this paper were implemented during the optimal route planning for service vehicles, which ensure the municipal waste collection of a particular commercial company.

Keywords: mathematical programming, *p*-median, GIS, routing problems, municipal waste collection.

JEL classification: C61
AMS classification: 90C05

1 Introduction

Vehicle route planning is an inherent part of decision making process of all subjects providing the transport network service. In practice, it may be cleaning and maintenance of roads, delivery of consignments to customers, municipal waste collection, separated municipal waste collection or other transport services. A general underlying problem is to design service vehicles routes while taking into account the chosen optimization criterion and also respecting all of the substantial constraints resulting from the real traffic. The most commonly used optimization criterion for designing service vehicles route is minimizing the total distance travelled (in kilometers). Basically, problems concerning to the service vehicle route designing can be divided into two groups. The first one represents such tasks, in which the subject of service is a road – *an edge* (snow cleaning, street cleaning etc.). The next group includes such problems that solve a point in a transport network – *a vertex* (customer, depot etc.). In this paper, we limit our focus to problems of the second set of tasks.

The initial task for vehicle route designing while the service of vertices in a transport network is taken into account, is the Traveling Salesman Problem requiring the most efficient Hamiltonian cycle. To determine the minimum Hamiltonian cycle (MHC), the following basic requirement must be fulfilled: the sum of requirements located at the vertices of a transport network is less than or equal to the service vehicle capacity. Nevertheless, this essential requirement is not usually met in practice. For this reason, it is necessary to consider not only a single ride, but a set of vehicle cycles. There is a whole range of methods for dealing with this problem, but we will primarily focus on the decomposition approaches. Firstly, the two-step decomposition Cluster-First Route-Second method can be used to generate clusters of vertices so that the sum of requirements in each cluster does not exceed the service vehicle capacity. Secondly, the Traveling Salesman Problem is subsequently solved in individual clusters using some of heuristic algorithms. A well-known Sweep algorithm can be used for clustering and this procedure is described in detail in [7]. In addition, it is also possible to use the Route-First Cluster-Second method. The Traveling Salesman Problem is solved by some heuristic method regardless of service vehicle capacity and only then the acquired Hamiltonian cycle is decomposed into individual circular rides taking into account the service

[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 Mining and Geology, Institute of Geoinformatics, 17. listopadu 15, Ostrava - Poruba, Czech republic, lucie.orlikova@vsb.cz
[3]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, marek.pomp@vsb.cz
[4]Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská 33, Ostrava 1, Czech republic, sarka.michalcova@vsb.cz

vehicle capacity. Both presented methods can be improved by using mathematical programming. The two-step approach is kept, but individual steps (both or at least one of steps) are solved exactly. An example of using mathematical programming for the Route-First Cluster-Second method is given in [5]. The presented contribution is devoted to the mathematical programming of the Cluster-First Route-Second method, namely using the *p*-median approach.

2 Formulation of problem

The real transport network is represented by the graph $N(V, E, l)$ with three attributes: $V = \{1, ..., m\}$ denotes a set of vertices, $E = \{1, ...n\}$ denotes a set of edges and l_{ij} represents an edge evaluation ij, where $i, j \in V$ is a road length (in kilometers). At the first vertex of the graph, there is a depot $\{d\}$ with service vehicles r, $r \in R$ and $R = \{1, ..., o\}$ denotes a set of service vehicles with the same capacity C. Customers j with requirements b_j are located at the vertices $j = 2, ..., m$ of the graph. At the same time, the following condition must be met: the sum of requirements of all customers is less than or equal to the sum of capacity of all service vehicles, hence:

$$\sum_{j=2}^{m} b_j \leq C \cdot o. \tag{1}$$

The main goal is to design individual vehicle cycles for vehicles r so that the total distance travelled while serving customers j is minimized. In addition, the following requirements must be met: each customer j is fully served by a single service vehicle driving r, the service vehicle capacity C is not exceeded and each vehicle r must be used at most once. A diagram of the directed graph N, which will also be used for the pattern Example 1, is given in Figure 1. The evaluation of edges denotes the road length (in kilometers) and the evaluation of vertices represents customer requirements.

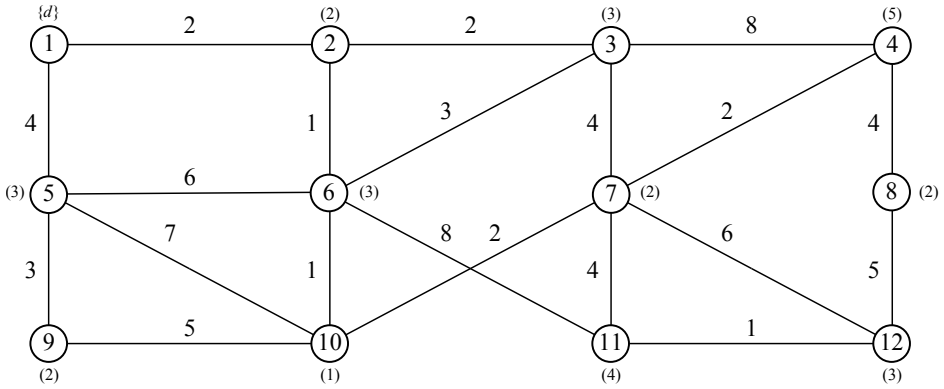


Figure 1 Diagram of the pattern graph N

The Cluster-First Route-Second method can be used to solve the problem as introduced above. This method is heuristic as a whole, but the first step can be exactly realized using the *p*-median model as a tool for clustering vertices $S_1, ..., S_q$, where $S_1 \cup, ..., \cup S_q \cup \{d\} = V$. The following condition must be fulfilled: the sum of customer requirements including individual clusters does not exceed the service vehicle capacity:

$$\sum_{j \in S_k} b_j \leq C, \text{ for } k = 1, ..., q. \tag{2}$$

Moreover, a further step is to find MHC using of some suitable approach, which depends on the task size (an exact resp. a heuristic approach).

3 The *p*-median problem

The *p*-median problem is a special case of the capacitated location-allocation problems and it can be used for clustering. The principle of this approach is following: two sets J, I are given; the set of customers $J = \{1, ..., n\}$ and the set of possible location of centres $I = \{1, ..., m\}$. Each customer j of the set J defines own service requirement (the size of this requirement is denoted as b_j). Each of centres, which will be located in some place of the set I, has the capacity K. A maximum predetermined number of the centres p may be located in the set I. The following condition must be met: the sum of capacities of all centres is greater than or equal to the sum of requirements of all customers:

$$\sum_{j=1}^{n} b_j \leq K \cdot p. \tag{3}$$

The distance between location of the centre i and the customer j can be modelled by the distance matrix d_{ij}. The goal is to design a location of p centres in a set of vertices I so that the sum of distances between the customer j and the nearest centre i is minimal and the capacity of any centre is not exceeded.

3.1 Mathematical model of the capacitated *p*-median problem

The allocation model of the described problem can be formulated as follows; [3] and [4]:

$$\text{Min} \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \cdot d_{ij} \tag{4}$$

$$\sum_{i=1}^{m} x_{ij} = 1, \text{ for } j = 1, ..., n \tag{5}$$

$$x_{ij} \leq y_i, \text{ for } i = 1, ..., m, \ j = 1, ..., n \tag{6}$$

$$\sum_{i=1}^{m} y_i \leq p \tag{7}$$

$$\sum_{j=1}^{n} x_{ij} \cdot b_j \leq K, \text{ for } i = 1, ..., m \tag{8}$$

$$x_{ij} \in \{0, 1\}, \text{ for } i = 1, ..., m, \ j = 1, ..., n \tag{9}$$

$$y_i \in \{0, 1\}, \text{ for } i = 1, ..., m \tag{10}$$

The bivalent variable x_{ij} represents an assignment of the customer j to the location i and in a similar way, the bivalent variable y_i represents an assignment of the centre in the location i. The objective function (4) gives the total distance travelled among customers and assigned centres. The conditions (5) ensure that each customer will be assigned to just one location. Furthermore, the conditions (6) express that each customer can be assigned only to a location with the assigned centre. The condition (7) means that only just p of centres will be located. What is more, the conditions (8) put the following constraint: the sum of customer requirements assigned to one centre does not exceed the specified capacity K. The obligatory conditions (9) and (10) define a domain of variables.

The above-mentioned model (4) – (10) can be used for already described clustering with following small adjustments. First of all, the set of customers J corresponds to the set of vertices of the graph V, from which the depot $\{d\}$ is removed, thus $J = V - \{d\}$. Moreover, it should be taken into account that the set of possible locations of centres is equal to the set of customers: $I = J$. In addition, the number of located centres p corresponds to the number of created clusters and the capacity of centres K corresponds to the sum of customer requirements, which are assigned to the individual clusters $S_1, ..., S_p$ (see the relationship (2) where $C = K$).

Example 1. The real transport network is presented by the vertex-edge graph N, which is shown in Figure 1. Individual customers $j \in J$ are located at the vertices of the graph $j = 2, ..., n$. The set of possible centre locations $i \in I$ corresponds to the set of customers, hence $I = J$. Customer requirements b_j are successively 2, 3, 5, 3, 3, 2, 2, 2, 1, 4, 3 units. What is more, the distance matrix d_{ij} is given for $i \in I$, $j \in J$ and the same capacity of centres $K = 10$ units. The goal is to design a location of three centres so that the sum of the distances from customer j to the nearest located center i is minimal and the capacity of any centre is not exceeded.

If we use the mathematical model (4) – (10) to solve the Example 1, we get both information about location of three centres and assigning customers to individual centres. These assignments can be considered as clusters. The solution is summarized in Table 1.

The centre at the vertex	Customer assignments	The sum of requirements [unit]
6	Cluster 1: {2, 5, 6, 9}	10
7	Cluster 2: {3, 4, 7}	10
12	Cluster 3: {8, 10, 11, 12}	10

Table 1 Customer assignment to the individual centres – clusters

As already mentioned, the customers assignment can be considered as clusters - i.e. subsets of customers, which need to be served from the depot $\{d\}$ located at the vertex 1. For this trivial task, it is easy to find three vehicle cycles starting and ending at the vertex 1. These vehicle cycles are presented in Table 2. Transit vertices in the framework of resulting vehicle cycles are completed with respect to the condition of the minimal way among served vertices (d_{ij} may be different from l_{ij}). Sequences of vertices in individual vehicle cycle can be easily projected to the graph N in Figure 1.

Vehicle cycle	Sequence of vertices	Transit vertices	Length [km]
1	1-**2**-**6**-10-**9**-**5**-1	10	16
2	1-2-**3**-**7**-**4**-**7**-10-6-2-1	2, 6, 10	18
3	1-2-6-**10**-7-4-**8**-**12**-**11**-7-**10**-6-2-1	2, 4, 6, 7	28

Table 2 Vehicle cycles in individual clusters

3.2 Mathematical model verification and validation

In the subchapter 3.1, the p-median mathematical model as a tool for clustering has been introduced and subsequently applied to the trivial Example 1. Furthermore, it is necessary to test the usage of this approach to the real-range tasks. For this purpose, the set of 20 tasks ranging 127 – 1389 vertices of the freely accessible database TSPLIB [9] was selected. Both the unit customer requirements and numbers of centres p with capacity K were determined so that the condition (6) was met. The Table 3 and Table 4 summarize the results of individual experiments.

Problem	1	2	3	4	5	6	7	8	9	10
Vertices	127	150	194	200	299	318	400	439	574	657
p	3	3	2	2	3	3	4	4	4	3
K	45	50	100	100	100	110	100	110	150	220
Time [sec]	8.7	7.4	6.9	5.0	150.4	24.6	50.7	131.1*	435.0	648.9

Table 3 Results of experiments 1 – 10

Problem	11	12	13	14	15	16	17	18	19	20
Vertices	734	783	813	929	980	1002	1060	1376	1379	1389
p	3	4	4	4	4	5	5	6	6	6
K	250	200	205	240	250	205	220	240	240	240
Time [sec]	305.7	149.27*	1506.7	439.6*	2005.4	290.6	502.5	4405.8	2744.4	1516.3*

Table 4 Results of experiments 11 – 20

While evaluating results of experiments presented in the Table 3 and Table 4, attention was especially focused on the computational time of the optimal solutions. Marked values * present the computational time of the tasks with acceptable solution where the gap is less than or equal to 20 %. Experiments show that the value of the computational time depending on the range of task increases exponentially. In the just described situation, the p-median approach was tested on the real task of a similar range and experiment's results are presented in the chapter 4.

4 Numerical experiments with the real data

Numerical experiments with the mathematical model (4) – (10) were implemented on the real problem of the municipal waste collection.

The specific commercial company ensures municipal waste collection in Bystřice nad Olší. There are 1496 customers with a unit service requirement on the territory of this municipality. At the present time, the service of this area is realized through three collection routes. The company requires redesign of three equal sets containing 500 customers approximately. These customers will be served in the framework of three vehicle cycles so that the total distance travelled is minimized. Customer's addresses and the service vehicle capacity are available and this capacity is sufficient for fulfilling of these 500 customer's needs. After geocoding of individual customer's addresses, the distance matrix was calculated (distances between customers and the depot) using the *ArcGIS* geoinformation software; [1] and [6].

The Cluster-First Route-Second method was used to solve the above mentioned problem. Firstly, clusters were created using the Sweep algorithm (SA), which was discussed in detail in [7]. Secondly, the *p*-median approach (*p*-med.) presented in this paper was used for generating clusters and the mathematical model of this approach was implemented in the computational environment *Xpress-IVE*; [10]. Finally, individual vehicle cycles, represented by the minimum Hamiltonian cycles (MHC) in clusters, were found using the heuristic algorithm in the computational environment *Wolfram Mathematica*; [2] and [8]. The achieved expected results are summarized in Table 5.

Cluster	Vertices (SA)	MHC [km]	Vertices (*p*-med.)	MHC [km]	Saving [km]
1	499	48.8	500	57.2	-8.4
2	499	54.1	496	44.8	9.3
3	498	57.0	500	37.7	19.3
SUMA	1496	159.9	1496	139.7	20.2

Table 5 Comparing experiment results to the real task

The Table 5 contains information about numbers of vertices in clusters, which were generated using the Sweep algorithm and the *p*-median approach. This table also provides information about the lengths of vehicle cycles searching in the individual clusters and achieved savings. If the Sweep algorithm is used for clustering, the total distance **159.9** km will be travelled while serving the territory. However, if the *p*-median approach is used for generating clusters, the total distance travelled is **139.7** km which means **20.2** km less. Let us recall that there were three circular rides of service vehicles in both cases.

5 Conclusion

The submitted contribution presented the usage of the *p*-median for the basic decomposition of the set of vertices to design service vehicle routing. The mathematical model of the *p*-median approach and its benefit of clustering was introduced in the subchapter 3.1. The subchapter 3.2 dealt with the description of experiments on the set of example tasks to verify the usability of the *p*-median model. Numerical experiment with the real data from municipal waste collection environment was given in the chapter 4. To summarize, taking into account the above mentioned *p*-median restrictions, this approach can be used for the problems of similar dimensions.

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. (2013). *GIS Tutorial 2: Spatial Analysis Workbook*. New York: Esri Press.

[2] Friedrich, V. (2013). *Matematika na počítači pro nematematiky*. Ostrava: Vydavatelství VŠB-TU Ostrava.

[3] Janáček, J. (2002). *Optimalizace na dopravních sítích*. Žilina: Žilinská univerzita.

[4] Janáček, J., Kvet, M. (2016). Sequential approximate approach to the *p*-median problem. *Computers & Industrial Engineering*, 94, 83–92.

[5] Kozel, P. (2017). Dekompozice okružních jízd s využitím matematického programování. In: *Perner's Contacts, Univerzita Pardubice, Dopravní fakulta Jana Pernera*, 3, 62–70.

[6] Ormsby, T., Napoleon, E. et al. (2010). *Getting to Know ArcGIS Desktop*. New York: Esri Press.

[7] Pomp, M., Kozel, P., Michalcová, Š., Orlíková, L. (2017). Using the Sweep Algorithm for decomposing a set of vertices and subsequent solution of the traveling salesman problem in decomposed subsets. In: *Proceed-*

ings of 35th International Conference Mathematical Methods in Economics 2017 (pp. 584–589). Uherské Hradiště.

[8] Wolfram, S. (1999). *The Mathematica Book*. New York: Cambridge University Press.

[9] *TSPLIB "Library of sample instances for the TSP (and related problems) from various sources and of various types"*. (2013). Heidelberg: Universität Heidelberg.

[10] *XPRESS-Mosel "User guide"*. (2005). United Kingdom: Dash Associates.

System Dynamics Model of Czech Regions' development

Igor Krejčí[1], Tomáš Šubrt[2], Jan Bartoška[3]

Abstract. Paper deals with the expression of the Czech regions as the dynamic complex systems. Such approach stresses the fact that regions and the regional actors are in interconnected, governed by feedbacks, causes and effects are nonlinear, moreover, the reaction is often delayed. As a result of dynamic complexity, the systems' behaviour (including the examined regions) is counterintuitive and policy resistant without using an appropriate tool.
For these purposes, we created the simulation model of Czech regions on the NUTS II level that explains the development of main regional attributes. The basic economic and demographic characteristics of the regions are specified in accordance with the official regional statistics. The missing model parameters were estimated by Powell optimisation minimising the difference between real and simulated data.
To express the attractiveness of the region with the impact on the internal migration and commuting to work, the basic economic indicators showed to be inadequate. For these reasons, the model includes generalised characteristics from the field of corporate social responsibility, which highly improve the accuracy of the simulation.

Keywords: Region, System dynamics, Computer simulation, Parametrisation.

JEL Classification: C44, C63, R11, R23
AMS Classification: 91B55, 90C30, 93C15

1 Introduction

The regional development could be described by many attributes, however, the analysis of development must reflect the dynamics. Looking at the region and its dynamics from the system perspective [11], [25], one must understand the consequences of dynamic complexity.

The term dynamic complexity expresses that the elements and agents in the system are tightly coupled and interdependent, the cause and effect aren't proportional, moreover, the effects are often delayed, the system contains feedbacks that strengthen or balance the actions [30], [31]. The common sense, the human capabilities to understand the behaviour of such system are very limited. The bounded rationality [26], [27], mental models [12] and cognitive limitations [3], [22] all lead to simplification of humans' understanding of complex systems. Therefore the complex systems are also characterised by counterintuitive behaviour and policy resistance [13], [30]. Consequently, many authors recommend the computer simulation as a tool that compensates the humans' limits and increases the understanding and predictability of the complex systems [4], [30].

In our paper, we focus on the development of Czech regions on the level NUTS 2. The aim is to implement the simulation model that would replicate the behaviour of the real system. We show the importance of the non-financial characteristics despite the parameters and variables of that part of the model aren't supported by data time series of high accuracy. Nevertheless, "to omit such variables is equivalent to saying they have zero effect – probably the only value that is known to be wrong." [15, p. 57]

At the end of the paper, we compare the results of the calibration with the model that doesn't incorporate the non-financial qualitative subsystem. The comparison shows the decrease of the accuracy of simulation without non-financial characteristics and that the behaviour of the inhabitants isn't possible to interpret without these attributes.

2 Material and Methods

The simulation model is created to characterize the development of the regions of the Czech Republic on the level NUTS2. The model is parametrized mainly on the basis of official demographics, regional statistics and statistics of national accounts from Czech Statistical Office [5–7]. The regional data on fixed capital were measured by

[1] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems of Engineering, Kamýcka 129, Prague 6 - Suchdol, Czech Republic, krejcii@pef.czu.cz
[2] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems of Engineering, Kamýcka 129, Prague 6 - Suchdol, Czech Republic, subrt@pef.czu.cz
[3] Czech University of Life Sciences Prague, Faculty of Economics and Management, Department of Systems of Engineering, Kamýcka 129, Prague 6 - Suchdol, Czech Republic, bartoska@pef.czu.cz

modified perpetual inventory method [18], [19]. All parameters and values that represent the value in currency (product, fixed capital, wage etc.) are evaluated at prices of the year 2010.

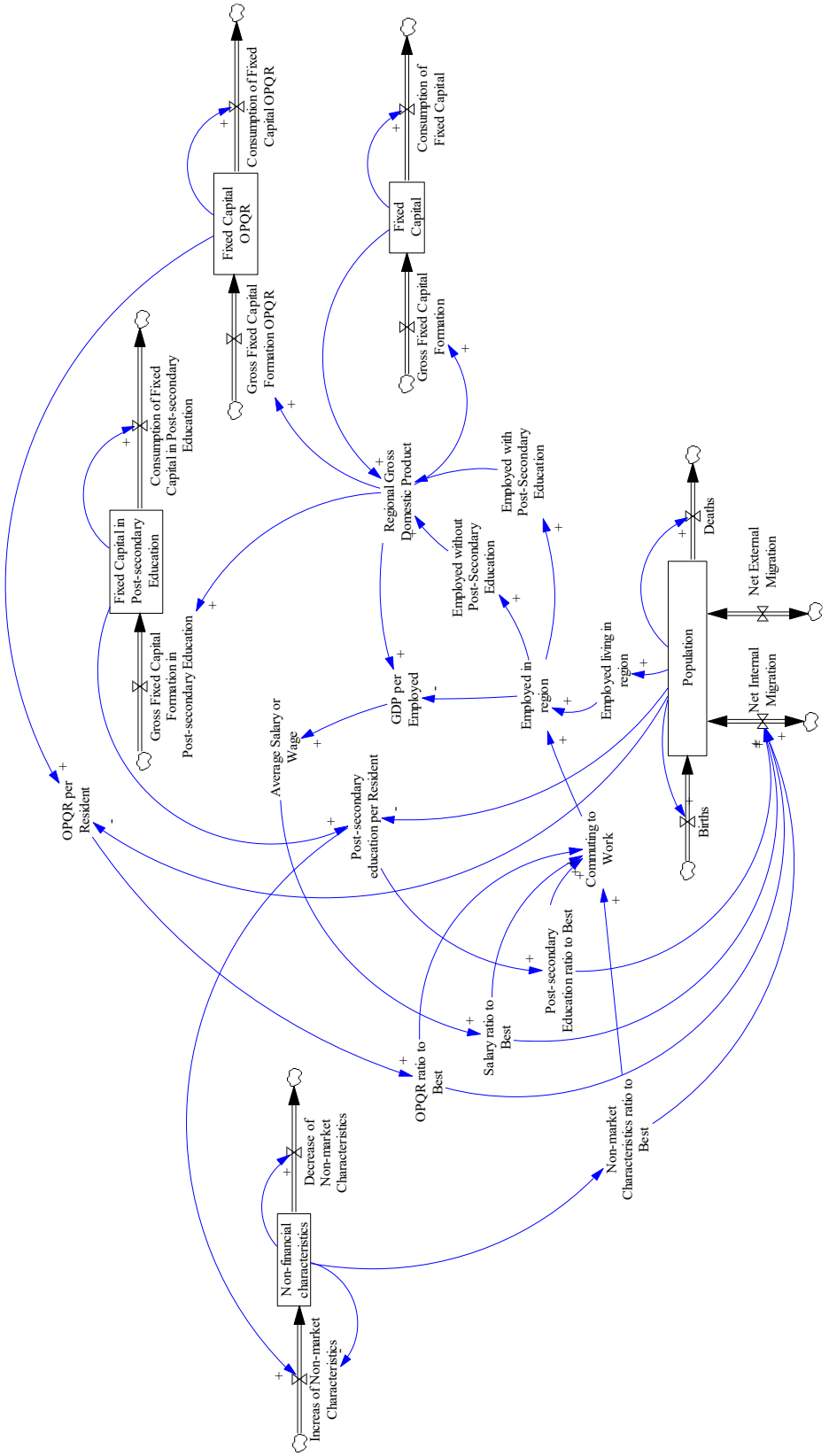


Figure 1 Simplified stock and flow diagram of dynamics of Czech regions

Figure 1 shows the simplified stock and flow diagram of each region. Due to the clarity reasons, many parameters are suppressed and the structures aggregated. Boxes represent the stock variables (accumulations) and

pipes denote the flow variables indicating inflows and outflows of the stocks. The boundary of the examined system is denoted by the clouds. For comparison, the full model that represents the interconnections of all 8 regions consists of 1127 variables and parameters.

The causal links' polarity has clear mathematical interpretation. The positive link polarity expresses that y increases above (decreases below) what it would have been if x increases (decreases) [30]:

$$\frac{\delta y}{\delta x} > 0 . \tag{1}$$

The negative link polarity denotes that y decreases below (increases above) what it would have been if x increases (decreases) [30]:

$$\frac{\delta y}{\delta x} < 0 . \tag{2}$$

Box variable represents stocks (levels) as the definite integral [30]:

$$s = \int_{T_0}^{T} (i - o)\mathrm{d}t + s_{T_0} . \tag{3}$$

The s represents a stock variable, i are all inflows, o are all outflows, T_0 is initial time, T is current time and t is any time between T and T_0.

For the output calculation (4) we extend the Cobb-Douglas productivity function [2], [28] similarly to [17], [20], which use human capital as additional production factor that contains the education. In our case, we parametrised the regional Cobb-Douglas production function that distinguishes the labour with and without post-secondary education:

$$y_t = a_t l_t^{\alpha} p_t^{\beta} k_t^{1-\alpha-\beta} , \tag{4}$$

where l is labour without post-secondary education, p represents labour with post-secondary education, k represents net fixed capital and a is total factor productivity. The parameters α and β represents output elasticities. The total factor productivity was smoothed by Hodrick-Prescott filter with λ=100 for annual data [1], [16].

The part that called "non-financial characteristics" in figure 1 expresses the fact that the motivation of inhabitants has various factors that influence their behaviour in regions. The model contains also other specified non-financial characteristics (post-secondary education, demography etc.), however, it was clear that it is necessary to add other characteristics that describe different aspects of regions' attractiveness. To express these additional characteristics we adopt the scales and Corporate Social Responsibility (CSR) areas from CSR reporting project [9]:

- Economic-employee area;
- Social-environmental area;
- Management of the non-financial characteristics (which is supposed to strengthen the impact of previous areas).

That structure is designed in a similar way as other types of capital, including service lives, however, because we use the scale 1-5 to measure the quality of these characteristics we implemented the balancing feedback loop that makes the increase of the value harder as the level of such characteristic grows.

The attractiveness of the region influence two important flows – the internal migration and commuting to work. The commuting to work is a function (5) of the wages w, economic-employee area e and the management of non-financial characteristics o. On the other hand, the internal migration (6) is influenced by many factors like amount of services s expressed as capital in industries OPQR according to classification CZ NACE [8], post-secondary education per inhabitant d and social-environmental area n. In most cases the factor was expressed as the ratio to the best value in the Czech Republic, however, the Central Bohemia, which is the ring around the capital, is characterised mostly by attributes of Prague (except the social-environmental area and management of the non-financial characteristics).

$$c_t = w_t^{\omega} e_t^{\varepsilon} o_t^{\mu} , \tag{5}$$

$$m_t = w_t^{\omega} e_t^{\varepsilon} m_t^{\mu} n_t^{\eta} s_t^{\sigma} d_t^{\delta} . \tag{6}$$

Table 1 provides the model boundary - the most significant endogenous, exogenous and excluded variables [23].

Endogenous variables	Exogenous variables	Excluded variables
Population	Total factor productivity	Expenditures on non-financial characteristics
Fixed capital stock	Net external migration	Subsidies
Gross fixed capital stock	Inflation	Transport accessibility
Commuting to work		
Net internal migration		
Wages		
Post-secondary education capital		

Table 1 Czech regions' dynamics – model boundary

Having the model structure of the non-financial characteristics we estimated the missing parameters (elasticities $\omega, \varepsilon, \mu, \eta, \sigma, \delta$ from (5) and (6) and service lives of non-financial characteristics) by Powell optimisation algorithm with multiple starts where we minimised the difference between simulated behaviour and time series of internal migration and commuting to work for period 2010-2016 [10], [24]. The model passed several tests - dimensional consistency test, integration error test, extreme condition test, parameter assessment etc. see [30] for details.

3 Results and discussion

The overall testing and model calibration took more than 300 million simulation runs. To evaluate the structure for the non-financial characteristics we also tested the calibration via Powell optimisation of the model without these characteristics. The table 2 shows Mean Absolute Percentage Error (MAPE), correlation coefficient (R) and coefficient of determination (R^2) between data series and model output (for discussion on these indicators for system dynamics models see [29]) and compare the best results of calibration for the model that includes or excludes the non-financial characteristics.

	Without non-financial structure			Including non-financial structure		
NUTS2	MAPE	R	R^2	MAPE	R	R^2
Internal emigration	0.065	0.386	0.149	0.035	0.846	0.715
Northwest CZ04	0.131	0.602	0.363	0.030	0.997	0.995
Northeast CZ05	0.205	0.443	0.196	0.034	0.983	0.966
Southeast CZ06	0.114	0.717	0.514	0.024	0.998	0.996
Central Moravia CZ07	0.095	0.387	0.150	0.026	0.947	0.897
Moravian-Silesian Region CZ08	0.065	0.386	0.149	0.035	0.846	0.715
Internal immigration						
Southwest CZ03	6.706	0.001	0.000	0.07	0.956	0.9144
Prague and Central Bohemia CZ01 and CZ 02	0.462	0.001	0.000	0.01	0.956	0.9135
Commuting to work						
Central Bohemia CZ02	0.462	0.915	0.837	0.017	0.962	0.925
Southwest CZ03	0.353	0.855	0.730	0.040	0.897	0.805
Northwest CZ04	0.428	0.598	0.358	0.065	0.871	0.759
Northeast CZ05	0.329	0.537	0.289	0.024	0.801	0.642
Southeast CZ06	0.129	0.908	0.824	0.116	0.915	0.836
Central Moravia CZ07	0.151	0.869	0.755	0.063	0.968	0.937
Moravian-Silesian Region CZ08	0.248	0.932	0.868	0.009	0.996	0.992

Table 2 Czech regions' dynamics – calibration results

The table 2 shows that the model could not be properly calibrated without the non-financial characteristics, especially the internal immigration was not possible to calibrate without additional characteristics. The results

confirm the importance of the system structure for its behaviour [14], [21]. Figure 2 illustrates the results of parameters' calibration on the net internal immigration to Prague and Central Bohemia.

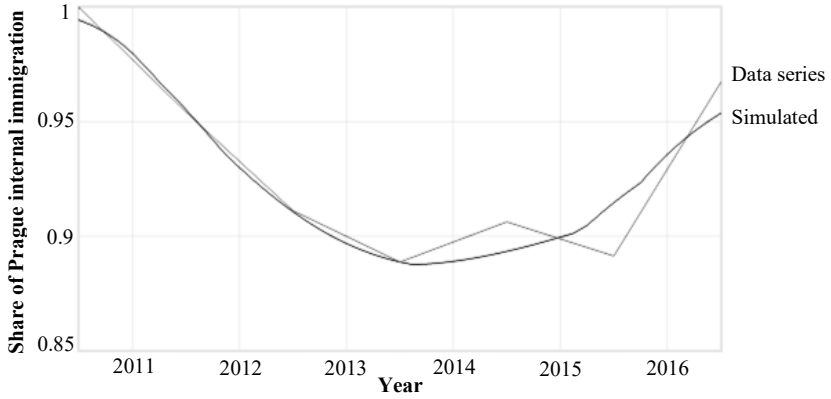


Figure 2 Share of Prague and Central Bohemia on internal immigration 2010-2016

4 Conclusion

The structure is the source of system's behaviour. The feedback loops, delays between cause and effect and the variability and organisation of the interconnections govern the evolution and development of the regions. In the current state, our simulation model could replicate the behaviour of the Czech regions. In this brief paper, we show the necessity of characteristics that aren't usually surveyed by statistical offices, however, improve the understanding of the system's behaviour and therefore also accuracy and precision of the simulation results.

Our future work will focus on the analysis of commuting to work in Prague as it could prove to be the problem with many consequences if the inflow of workers will significantly increase. For this analysis, it will be necessary to identify leverage points where the small change leads to biggest changes. Nevertheless, it will be also necessary to express that change in economic terms. In current state of art, we excluded the price of the increase of the non-financial characteristics, however, for proper policy design that part will be deeply investigated.

Acknowledgements

The article is supported by the grant project No. TD03000146 of the Technology Agency of the Czech Republic

References

[1] Brandner, P. Diebalek, L. & Schuberth, H. (1998). *Structural Budget Deficits and Sustainability of Fiscal Positions in the European Union*. Working Paper, No. 26. Wien: Oesterreichische Nationalbank
[2] Čadil, J., Vltavská, K., Krejčí, I., Hartman D. & Brabec, M. (2017). Aggregate Production Function and Income Identity – Empirical Analysis. *International Journal of Economic Sciences, 6*, 1-17.
[3] Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.
[4] Coyle, R. G. (1996). *System dynamics modelling: a practical approach*. London: Chapman & Hall/CRC
[5] Czech Statistical Office (2017). *Czech Demographic Handbook - 2016* [Online]. Available: https://www.czso.cz/csu/czso/czech-demographic-handbook-2016 [26 Nov 2017].
[6] Czech Statistical Office (2018). *National Accounts* [Online]. Available: http://apl.czso.cz/pll/rocenka/rocenka.indexnu?mylang=en [18 Feb 2018].
[7] Czech Statistical Office (2018). *Regional Accounts* [Online]. Available: http://apl.czso.cz/pll/rocenka/rocenka.indexnu_reg?mylang=en [20 Feb 2018].
[8] Czech Statistical Office (2018). *Classifications* [Online]. Available: https://www.czso.cz/csu/czso/classifications [28 Feb 2018].
[9] Czech University of Life Sciences Prague, University of Economics Prague & Charles University (2017). *CSR Reporting* [Online]. Available: https://csr-reporting.czu.cz/ [11 Nov 2017]
[10] Dangerfield, B. & Roberts C. (1999) Optimisation as a statistical estimation tool: An example in estimating the AIDS treatment-free incubation period distribution. *System Dynamics Review*, *15*, 273–291.
[11] Desthieux, G., Joerin F. & Lebreton, M. (2010). Ulysse: a qualitative tool for eliciting mental models of complex systems: Methodological approach and application to regional development in Atlantic Canada. *System Dynamics Review, 26*, 163–192.
[12] Doyle, J. K. & Ford, D. N. (1998). Mental models concepts for system dynamics research. *System Dynamics Review, 14*, 3–29.

[13] Forrester, J. W. (1971). Counterintuitive behavior of social systems. *Technological Forecasting and Social Change, 3*, 1–22.
[14] Forrester, J. W. (1987). 14 “obvious Truths". *System Dynamics Review, 3*, 156–159.
[15] Forrester, J. W. (1961). *Industrial dynamics*. Cambridge, Mass.: M.I.T. Press.
[16] Giorno, C., Richardson, P. Roseveare, D. & Van Den Noord, P. (1995). Potential Output, Output Gaps And Structural Budget Balances. *OECD Journal: Economic Studies, 24*, 167-209.
[17] Jiménez Bandala, C. A. & Andrade, L. A. (2017). Education, Poverty and the Trap of Poor Countries in the Face of Development. *Journal on Efficiency and Responsibility in Education and Science, 10*, 101–108.
[18] Krejčí, I., Mazouch, P., Vltavská, K. & Kvasnička, R. (2015). Age of machinery and equipment in the Czech agriculture. *Agricultural Economics, 61*, 356-366.
[19] Krejčí, I. & Sixta, J. (2012). Využití alternativních metod při odhadech stavů a spotřeby fixního kapitálu. *Politická ekonomie, 60*, 780–800.
[20] Mankiw, N. G., Romer, D. & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics, 107*, 407–437.
[21] Meadows, D. H. & Wright, D. (2008). *Thinking in systems: a primer.* White River Junction, Vt.: Chelsea Green Pub.
[22] Miller, G. A. (1956). The Magical Number 7, Plus Or Minus 2 - some Limits on our Capacity for Processing Information. *Psychological review, 63*, 81–97.
[23] Pierson, K. & Sterman, J. D. (2013). Cyclical dynamics of airline industry earnings. *System Dynamics Review, 29*, 129–156.
[24] Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B.P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. *2nd ed*. New York, NY: Cambridge University Press,.
[25] Rich, K. M. (2008). An interregional system dynamics model of animal disease control: applications to foot-and-mouth disease in the Southern Cone of South America. *System Dynamics Review, 24*, 67–96.
[26] Simon, H A. (1956). Rational Choice and the Structure of the Environment. *Psychological review, 63*, 129–138.
[27] Simon, H A. (1979). Rational Decision-Making in Business Organizations. *American Economic Review*, 69*,* 493–513.
[28] Solow, R. M. (1957). Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics, 39*, 312-320.
[29] Sterman, J. D. (1984). Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica, 10*, 51-66.
[30] Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin/McGraw-Hill.
[31] System Dynamics Society (2018). *What is System Dynamics* [online]. Available: https://www.systemdynamics.org/what-is-sd [19 Mar 2018]

Analysis of the Slovak economy using VAR models

Radmila Krkošková[1]

Abstract. The paper aims to analyse and model chose macroeconomic variables of the Slovak economy and their dynamics using VAR models. This article shows the application of selected model on real-time series of chosen macroeconomic indicators using four variables (R – interest rate, M – money supply (M2), P – price level (CPI), Y – GDP). We identify and test four long-run relationships. There are described Granger causality, impulse response function, cointegration and error correction models. The estimation outputs are interpreted. The quality econometric model for modelling macroeconomic time series in Slovak economy is discussed. The calculations used EViews software version 9. The structural model is estimated for the Slovak economy. The data used have the character of a quarterly time series in the period from Q1/2005 to Q4/2017. The data source was the National Bank of Slovakia and the Eurostat database.

Keywords: ADF test of stationarity, Cointegration relation, Granger causality, VAR model, VEC model.

JEL Classification: C10, E27

AMS Classification: 62P20, 91B64

1 Theoretical Background and Data

The aim of the paper is macroeconomic analysis of the Slovak economy based on the application of vector autoregression models (VAR models). VAR models are widely used in quantifying and analysing macroeconomic relationships, in compiling economic policies, and in analysing and anticipating their impacts on economic development. In addition to VAR models, other econometric tools such as Granger causality, response function analysis and time series cointegration are listed in the paper.

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 [9], 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 [3]. Garrat et al. [5] 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. 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. Paper [2] is focused on the Swiss economy in the period from Q1/1976 to Q4/2006. Papers [6], [7] or [12] are focused on the Czech economy in the period from Q1/1995 to Q4/2009, or from Q1/2005 to Q4/2016. These articles also confirm the presence of five cointegration relationships.

A VAR model which includes the following variables is quite often used in applied macroeconomics: IR (interest rate), M2 (money supply), CPI (consumer price index) and GDP (gross domestic product). The analysis was performed on quarterly data for the period from 1Q/2005 to 4Q/2017. The data source was the National Bank of Slovakia and the Eurostat database. Interest rate (IR) is expressed as a percentage, consumer price index (CPI) is expressed in base indices (2005 = 100), money supply is in the form of monetary aggregate (M2)

[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: krkoskova@opf.slu.cz.

in millions of EUR, and real gross domestic product (GDP) in millions of EUR. All variables were seasonally cleaned using the X12 ARIMA method and were logarithmized.

Based on the analysis, we will find out if the results match the economic theory and answer these questions:

1. Does the monetary aggregate affect the gross domestic product and the consumer price index development?
2. Is there a direct relationship between the interest rate and the consumer price index?
3. Does the interest rate affect other variables?
4. Which of these variables form short-term relationships?

2 Model for the Slovak economy

The second chapter deals with the estimation and testing of the basic macroeconomic model for the Slovak economy in the period from Q1/2005 to Q4/2017. 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 Granger causality. As the third step, the impulse-response analysis is used. As the fourth step, the cointegration relationships for the VAR(p) model are tested using the Johansen's method and a number of cointegration relationships is determined. This subchapter provides the estimations of cointegration vectors.

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
L_CPI	c	– 2.69	0.08	N	*D*(*L_CPI*)	c	– 4.59	0.0005	S
L_GDP	c	– 2.77	0.069	N	*D*(*L_GDP*)	c	– 4.88	0.0002	S
L_IR	c	0.91	0.99	N	*D*(*L_IR*)	c	– 4.42	0.0008	S
L_M2	c	– 2.04	0.26	N	*D*(*L_M2*)	c	– 6.33	0.0000	S

Table 1 Testing the unit root of the variables in levels and their first differences

The variables (*L_CPI; L_GDP; L_IR; L_M2*) 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, Granger causality tests

This chapter describes the estimation of VAR(*p*) model. Model estimation includes a determination of *p* order for delayed variables (*L_CPI; L_GDP; L_IR; L_M2*) 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	181.3067	NA	7.27e-09	–7.387780	–7.231847	–7.328853
1	469.6192	516.5598*	8.62e-14*	–18.73413*	–17.95446*	–18.43949*
2	480.4386	17.58154	1.09e-13	–18.51827	–17.11487	–17.98793
3	497.4264	24.77394	1.08e-13	–18.55943	–16.53230	–17.79338
4	516.0267	24.02537	1.04e-13	–18.66778	–16.01691	–17.66601

* indicates lag order selected by the criterion

Table 2 VAR lag order selection criteria

A four-equation unlimited VAR model was estimated. The optimal length of delay was selected according to the values of information criteria listed in Table 2. The values indicate inclusion of only one delay. The VAR(1) model estimates are given in Table 3, standard errors in ( ), t-statistics in [ ].

	L_CPI	L_GDP	L_IR	L_M2
L_CPI(–1)	0.926099***	0.002462	–0.796669	0.014685
	(0.03513)	(0.10655)	(1.25162)	(0.10739)
	[26.3591]	[0.02311]	[–0.63651]	[0.13674]
L_GDP(–1)	0.066975***	1.082115***	1.999529**	0.295743***
	(0.02362)	(0.07163)	(0.84140)	(0.07219)
	[2.83565]	[15.1080]	[2.37642]	[4.09665]
L_IR(–1)	0.000147	–0.002748	0.951856***	–0.012462***
	(0.00118)	(0.00358)	(0.04204)	(0.00361)
	[0.12489]	[–0.76795]	[22.6404]	[–3.45474]
L_M2(–1)	–0.026171	–0.075943	–1.478223*	0.721521***
	(0.02155)	(0.06536)	(0.76779)	(0.06588)
	[–1.21430]	[–1.16193]	[–1.92529]	[10.9528]
R-squared	0.994507	0.984890	0.989894	0.995194

Statistical significance at the 0.01 level (***),at the 0.05 level (**),at the 0.1 level (*)

Table 3 Estimation of VAR(1) model

The CPI price level is affected by the value in the first delay and the GDP variable. The GDP variable is considerably persistent and only correlated with its value in the first delay. The IR interest rate equation shows that IR is affected by all model variables, except for the CPI variable. The monetary aggregate is affected by all model variables, except for the CPI variable, at 1% materiality level.

Granger causality, which deals with testing of short-term relationships, was also examined for the estimated VAR(1) model. The relationships in which Granger causality was demonstrated are shown in Table 4. Mutual Granger causality of D(L_M2) and D(L_IR) variables was demonstrated at 5% materiality level. Furthermore, the influence of D(L_GDP) on D(L_IR), D(L_M2), D(L_CPI) was demonstrated at 5% materiality level. The gross domestic product time series affects the interest rate, the monetary aggregate and the consumer price index within the meaning of Granger causality. Other short-term relationships were not identified.

Null Hypothesis	**Reject H_0**
D(L_M2) does not Granger Cause D(L_IR)	YES
D(L_IR) does not Granger Cause (L_M2)	YES
D(L_GDP) does not Granger Cause D(L_IR)	YES
D(L_GDP) does not Granger Cause D(L_M2)	YES
D(L_GDP) does not Granger Cause D(L_CPI)	YES

Table 4 Granger Causality

The diagnostic test was performed to assess the quality of the VAR(1). These tests include the conditions of stationarity, non-correlatability and normality test of residual component. VAR(1) 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 VAR(1) model is stationary. Non-correlatability of residual component of the estimated VAR(1) 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 Impulse-response analysis

Impulse-response analysis allows analysis of both the short-term and long-term relations between the analysed variables based on the derived model. Arlt [1] states that the impulse-response analysis is related to the question of what reaction in one time series will be caused by an impulse in another time series within a system that contains multiple time series. This is the study of the relation between two one-dimensional time series in a multi-dimensional system.

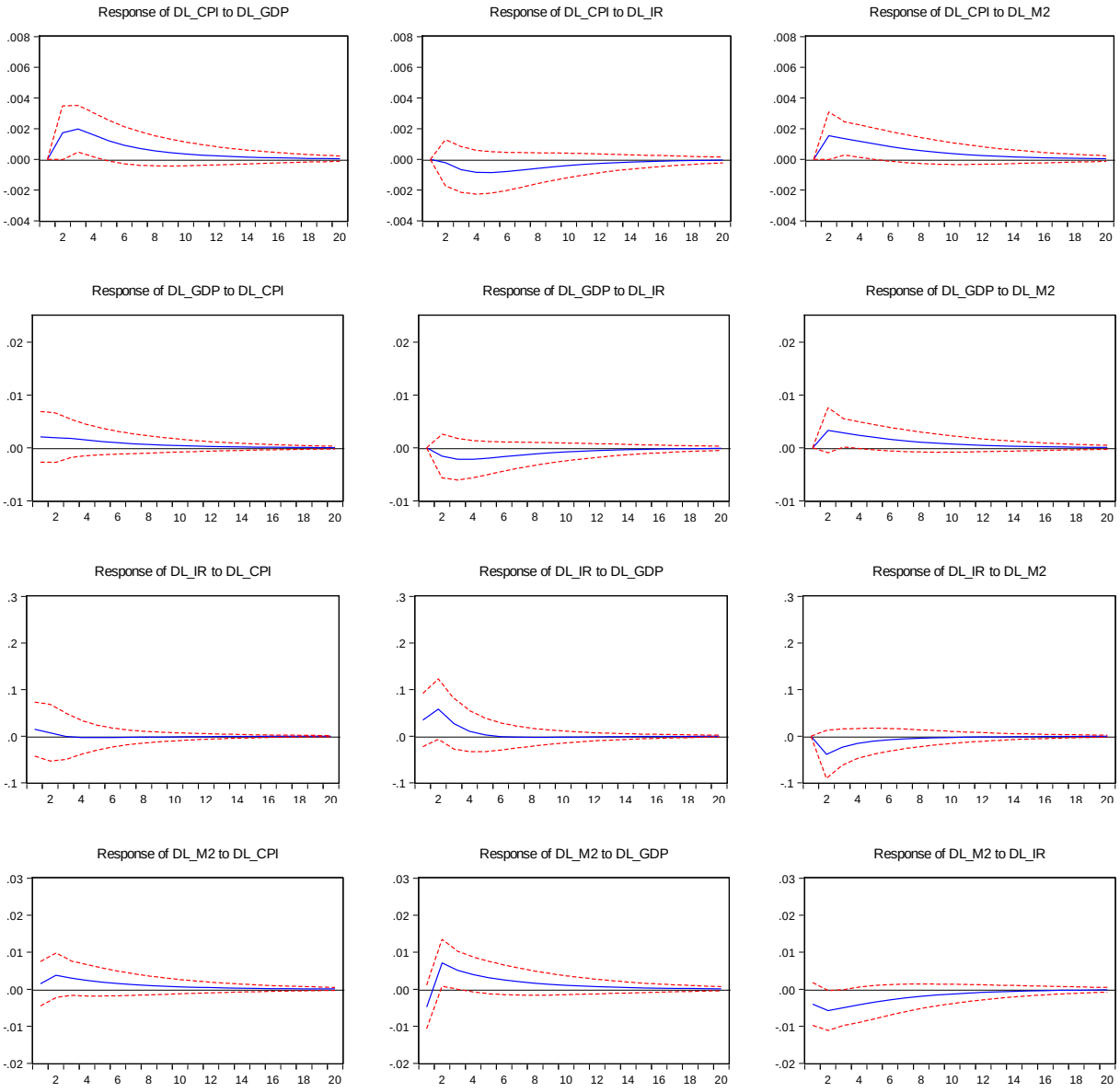


Figure 1 Response to Cholesky One S.S.Innnovations

The M2 unit exogenous shock triggers a GDP growth of approximately 3 years. Growth rate is increasing in the first half-year, then decreasing, and the effects of expansionary stimulus of monetary policy are gradually dwindling and GDP returns to equilibrium. This is in line with Keynesian theory that monetary policy can affect the real product. The impulse-response analysis suggests that money is not neutral in the short term; the GDP response to the exogenous shock in the M2 monetary aggregate is relatively strong and fades away for a long time. A similar reaction is triggered by the unit exogenous shock in M2 also in the consumer price index (CPI) and interest rate (IR). Positive unit exogenous shock in GDP triggers an increase in the monetary aggregate growth rate, culminating in the 3rd period and the growth rate is slowly decreasing. The whole process takes about 4 years. The unit exogenous shock in the price level then causes an increase in the monetary aggregate, which peaked in the second period, and then the growth rate slowly declines (but is still positive) until it returns to balance after 4 years. We will also look at the effects of unit exogenous shock in interest rate on other endogenous variables. The monetary aggregate reacts with a slight decline, with a subsequent return to the equilibrium level, the reaction being very weak. GDP is growing, the growth rate is decreasing after two periods, but it is positive, until the effect fades away (approximately after two years).

2.4 Cointegration relationship testing, VECM

This chapter deals with the testing of the number of cointegration relationships in VAR(1) model for the endogenous variables (*L_CPI; L_GDP; L_IR; L_M2*) using the Johansen's method, as shown in [4], [10]. Table 5 confirms the existence of 2 cointegration relationships for VECM(1). 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 *	0.506811	71.49010	54.07904	0.0007
At most 1 *	0.305614	36.14694	35.19275	0.0393
At most 2	0.237229	17.91060	20.26184	0.1021
At most 3	0.083703	4.370746	9.164546	0.3597

Trace test indicates 2 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level

Table 5 Unrestricted Cointegration Rank Test (Trace)

Restrictions were required to estimate two cointegration vectors. Restrictions were introduced, stemming from long-term demand for money and the Fisher effect [1]. The M2 monetary aggregate was set to 1 and IR to 0 in the first cointegration vector. GDP and M2 were set to 0 and interest rate to 1 in the second cointegration vector. Following the introduction of the restrictions, the following estimates of cointegration vectors were obtained.

$$C_{1t} = LM2_t - 1.31 \cdot LGDP_t - 1.43 \cdot LCPI_t \quad (1)$$

$$C_{2t} = LIR_t - 1.12 \cdot LCPI_t \quad (2)$$

According to relationship (1), the monetary aggregate M2 is directly dependent on GDP and CPI. Relationship (2) shows that the interest rate is directly dependent on CPI. These results are in line with economic theory.

Non-correlatability of residual component of the estimated VECM(1) 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. The residual component homoscedasticity null hypothesis was not rejected at the 5% significance level, as the results show Chi-sq = 92.84, df = 80; Prob. = 0.1545. The test "No Cross Terms" was performed (only levels and squares).

Table 6 shows the predictive ability of each estimated equation in VECM(1) with restrictions. The average level of adj.R^2 = 0.666 and it ranges within the interval from 0.541 (CPI equation) to 0.784 (M2 equation). These explanation measures compare favourably with the English [5] and Swiss study [2].

	D(L_CPI)	D(L_GDP)	D(L_IR)	D(L_M2)
R^2	0.695	0.758	0.766	0.806
Adj.R^2	0.541	0.651	0.688	0.784

Table 6 Determination coefficients for the estimated equations of the resulting model

Finally, we can say that the estimation of the VECM(1) 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.

3 Conclusion

This article dealt with the modelling of long-run structural equilibrium relationships for the Slovak economy in the period from 2005 to 2017. Modelling using cointegration vector autoregressive model with restrictions for endogenous variables (*L_CPI; L_GDP; L_IR; L_M2*) was used. Estimation of the VECM(1) model with re-

strictions 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.

The model with a single delay was selected as the most appropriate model based on the selected criteria. One delay indicates that dynamics does not play a material role in the relationships between the selected variables as assumed. The answers to the first two questions are given by two cointegration relationships. According to relationship (1), the monetary aggregate M2 is directly dependent on GDP and CPI. Relationship (2) shows that the interest rate is directly dependent on CPI. These results are in line with economic theory. The answer to the third question, whether the interest rate affects the other variables, gives an impulse response analysis. The monetary aggregate reacts with a slight decline, with a subsequent return to the equilibrium level, the reaction being very weak. GDP is growing, the growth rate is decreasing after two periods, but it is positive, until the effect fades away. CPI response to IR change is very weak. The last question concerned the existence of long-term relationships. These short-term relationships were identified: the relationship between M2 and IR, as well as these causal relations GDP→IR, GDP→CPI, GDP→M2.

The disadvantage of applying VAR models to macroeconomic aggregates is the necessity of stationarity of used time series. Stationarization causes a loss of certain part of information. Another disadvantage is a rapid increase in the estimated parameters with the inclusion of a large number of endogenous variables. It is usually necessary to select VAR models with a low length of maximum delay for quarterly macroeconomic data, which does not allow sufficiently capturing the model dynamics. Despite these shortcomings, the concept of VAR models appears to be appropriate in the area of macroeconomic analysis of the Slovak economy.

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 Slovak economy underwent a currency crisis characterized by atypical behaviour of interest rates and monetary indicators. 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, and extend the model of time series: the exchange rate SKK/EUR, industrial producer prices and foreign annual interest rate EURIBOR, as specified in [8].

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 2018.

References

[1] Arlt, T., and Arltová, M. (2009). *Ekonomické časové řady.* Professional Publishing, Praha.
[2] Assenmacher-Weschke,K., and Pesaran,M.H. (2008). A VECX Model of the Swiss Economy. *CESifo Working Paper,*No. 2281. University of Cambridge, Cambridge.
[3] Blanchard, O. J., and Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbanes. *American Economic Review* **79**(4), 655-673.
[4] Enders, W. (2001). Cointegration and Threshold Adjustment. *Journal of Business & Economic Statistics* **19**(2), 166-176.
[5] Garratt, A., and Lee,K. (2003). A Long Run Structural Macroeconometric Model of the UK. (Pesaran, M.H., and Shin, Y., eds.). *The Economic Journal* **113**(487), 412-455.
[6] Hančlová, J. (2009). Kvartální makroekonomický model české ekonomiky. In: *Aktuální aspekty české a světové ekonomiky,* Technická univerzita Liberec, Liberec, 138-147.
[7] Hančlová, J., and Kubicová, I. (2010). *Makroekonometreické modelování české ekonomiky a vybraných ekonomik EU.* (Macháček,M.,Melecký, L.,Nevima,J. and Ramík,J.,eds.).VŠB-TU,Ostrava, 31 – 88.
[8] Hančlová, J., and Juriová, J.(2017). Long-run Relations in a Small Open Economy of the Czech Republic and the Slovak Republic. (Závacká,J.,eds.) *Ekonomický časopis*. Slovenská akademie věd, Ekonomický ústav, **65**(5), 408-425.
[9] Hoover,K.D., and Johansen,S. (2008). Allowing the Data to Speak Freely: The Macroeconometrics of the Cointegrated Vector Autoregression. (Juselius,K., eds.).*American Economic Review:Paper and Proceedings* **98**(2), 251-255.
[10] Johansen, S.(1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.
[11] Schneider, E., and Chen,P. (2007). A Long-Run Structural Macroeconometric Model for Germany. (Frohn,J., eds.). *Economics Discussion Papers: The Open-Access, Open-Assessment E-Journal,*No.47.
[12] Stoklasová, R. (2017). Macroeconomic Modelling using Cointegration Vector Autoregression. In: *35th International Conference on Mathematical Methods in Economics*, University of Hradec Králové, 732-738.

Financial Cycle and Business Cycle Interactions in Countries with Quantitative Easing

Zuzana Kučerová[1,2], Jitka Poměnková[1], Eva Klejmová[1]

Abstract. Generally, financial crises lead to large macroeconomic disturbances in the world economy. However, several central banks adopted the unconventional monetary policies after the financial crisis of 2007 (quantitative easing, negative interest rates, forex interventions, etc.). In our paper, we focus on the interrelationship between the financial and the macroeconomic sector, i.e. the financial and business cycles, of the selected countries with the regime of quantitative easing (i.e. the euro area, Japan, the United Kingdom and the United States) to find whether these economies face the same co-movement of financial and business cycles. As such, the aim of the paper is to examine the mutual co-movement between the credit supply growth and real GDP growth for each of the selected countries. For this assessment, we use the wavelet power cross-spectrum to estimate spectral features of co-moved time series as a function of time. The significant co-moved part is identified via testing. In this way, we identify regions with periodicities of co-movement that evolved over time with the most extensive co-movoment of financial and business cycles in the US economy and the least extensive co-movement in the euro area economy.

Keywords: power wavelet co-spectrum, co-movement, bank loans, credit standards.

JEL Classification: C19, C63, E32
AMS Classification: 62P20, 62M15

1 Introduction

The financial crisis of 2007 revealed that financial market supervision and regulation had been quite weak in some areas of financial market. Because financial cycles differ from business cycles and have an impact on business cycles and economic recoveries (see Claessens et al. [5]), traditional macroeconomic policy is not able to address them and treat them properly. Therefore, the crisis changed the character of traditional macroeconomic policy (mainly fiscal and monetary) and macro-prudential policy is currently a necessary part of macroeconomic policy of each state. According to Drehman et al. [7], the length and amplitude of the financial cycle have increased since the early 1980s, partly as a result of financial liberalisation and innovations and less tight monetary policy in this period (compared to previously applied Keynesian policies). As such, it is of high importance to identify the financial cycles and show how they relate to business cycles, so policy makers can adjust the policy measures not only to business cycles but also to financial cycles.

A measure of co-movement between time series can be done by several approaches. The popular one in the last decade especially among economist is wavelet approach in the time-frequency domain. There are also other time-frequency techniques, such as Short Term Fourier Transform, Time-Frequency Autoregressive Process, but they are more popular for engineering due to character of engineering data. The economic data are moreover non-stationary ad are the composition of several frequency component with unique complicated patterns over time (Ftiti et al. [9]). Thus, wavelet analysis is suitable instrument which can capture the cyclical structure in time having good time resolution (Aloui et al. [1]). There are panoply of economic applications, such as (Ftiti [9], Berdiev *et al*. [4], Fidrmuc *et al*. [8], Aquiraria-Conraria et al. [2], Rua [15], Tiwari [18] and many others) focuses of economic time series synchrony. Rua [15] investigate the relationship between money growth and inflation in the euro area using wavelets and found out the stronger link between inflation and money growth at low frequencies over the whole sample period. Aloui et al. [1] focuses on business cycle synchronisation of Gulf

[1] Brno University of Technology, Faculty of Electrical Engineering and Communication, Technická 12, 616 00 Brno, Czech Republic, e-mail: pomenkaj@feec.vutbr.cz, xklejm00@stud.feec.vutbr.cz.
[2] Mendel University in Brno, Faculty of Business and Administration, Zemědělská 1665/1, 613 00 Brno, Czech Republic, e-mail: zuzana.kucerova@mendelu.cz.

Cooperation Countries via wavelets and uncover that the real growth rates in most of the countries within the GCC region commove with the others over the short and medium terms, while long-term co-movement of real growth rates is only found in seven out of the 15 country pairs. Tiwari et al. [18] assess the level of co-movements between the stock markets of the PIIGS, UK and Germany. They found via wavelets significant co-movement in the short-run term during financial distress episodes, while in the long-run for the entire analysed horizon. Furthermore, at low-frequency levels, the PIIGS stock markets are more synchronised with Germany than with the UK, while at high-frequency component an opposite result is obtained. Berdiev et al. [4] analyse synchronization of growth cycles between China, Japan, the United States and other Asia-Pacific countries using wavelet analysis. Their results shows that the growth cycles of China, Japan, and the United States are synchronized with the other Asia- Pacific economies, the strength of business cycle synchronization fluctuates across frequencies and over time. Fidrmuc et al. [8] study globalization and co-movement in business cycles in China and G7 countries and found out significant relationship between time-varying wavelet measure of co-movement and bilateral trade only for business cycle frequencies.

In many economic papers, financial cycles are measured by the volume of credit provided to economy and pay special attention to medium-term frequencies (in order to demonstrate the effect of the financial crisis). Galati et al. [10] measure the financial cycle using the credit-to-GDP ratio and house prices for the United States and the five largest euro area countries and find that medium-term financial cycles are longer than business cycles and identify differences among analysed countries. Drehmann et al. [7] use a set of variables (credit-to-GDP ratio, equity prices, house prices, asset prices) and focus on credit and property cycles in Australia, Germany, Japan, Norway, Sweden, the United Kingdom and the United States)and state that financial cycles are considerably longer than business cycles and that financial crises cause deeper business cycle recessions. In this context, one should focus on medium-term, not on business cycle, frequencies (cycles of 16 years or longer). Moreover, peaks of financial cycles appear together with financial crises. Using the multivariate spectral method of deriving common cyclical frequencies, Schüler et al. [17] focus on credit (house) and asset (equity and bonds) markets in selected European Union countries and whole euro area. They identify a common movement in the analysed area but also the existence of heterogeneity in the character of financial cycles across analysed countries. Therefore, they argue that the EU macro-prudential policy is necessary, but it should take it into account when formulating general policy measures. Claessens et al. [5] analyse credit (house) and equity markets in a large set of 44 countries and conclude that financial and business cycles are strongly interlinked and that economic recessions caused by financial crises are more pronounced than other recessions. Moreover, economic recovery is longer after the crises at asset markets compared to credit markets.

A limited attention has been paid to co-movements between financial and business cycles so far. In this area of research, we can mention a paper by Rünstler and Vlekke [16] who use multivariate unobserved components models to estimate the cyclical components in credit volumes, house prices and GDP in the US and the five largest European economies. They identify long financial cycles and the existence of correlation with the medium-term business cycles (longer than common 8-32 quarters used for the definition of business cycle) and some country differences thanks to national housing market characteristics. Igan et al. [13] measure the co-movement of house, credit and business cycles in the US economy and conclude that there are some institutional differences among analysed countries that have an impact on the character of cycles in the medium-term period. However, house prices lead credit and real activity in the long-term perspective.

In our paper, we use credit volumes as a key financial cycle variable (as such, we focus on credit cycles) to examine how the financial cycles are synchronised with the business cycle of a selected country in a set of countries which adopted the regime of quantitative easing. The aim of the paper is to measure the co-movement of financial and business cycles in euro area, Japan, the UK and the US to find whether these economies face the same co-movement of financial and business cycles. However, we neither address the causal relationships between financial and GDP cycles nor try to assess the impact of quantitative easing on the co-movement of financial and business cycles in our paper. Therefore, we do not focus on the specific period with the unconventional monetary policy regime. Instead, we apply the wavelet approach in the time-frequency domain on economic data and show whether there are some similarities or differences among the analysed countries.

2 Data and indicators

We use seasonally adjusted quarterly data of real gross domestic product (*GDP*), volume index in OECD reference year 2010 (OECD [14]), and unadjusted quarterly data of the total credit to private non-financial sector (BIS [3]) of the United Kingdom (UK) and the United States (US) in 1963/Q2-2017/Q3 (n=2018), for Japan in 1994/Q1-2017/Q3 (n=95) and for Euro Area (EA) in 1999/Q1-2017/Q3 (n=75). For each country, we analyse the co-movement of the real GDP with the total credit (*Credit*). All variables are in the first differences of logarithms. The data of the variable Credit are seasonally adjusted using TRAMO/SEATS. Given the quarterly

character of the data, we can denote the sampling frequency f_s=4 samples per year. The representations of time series used in the analysis are depicted in Figure 1.

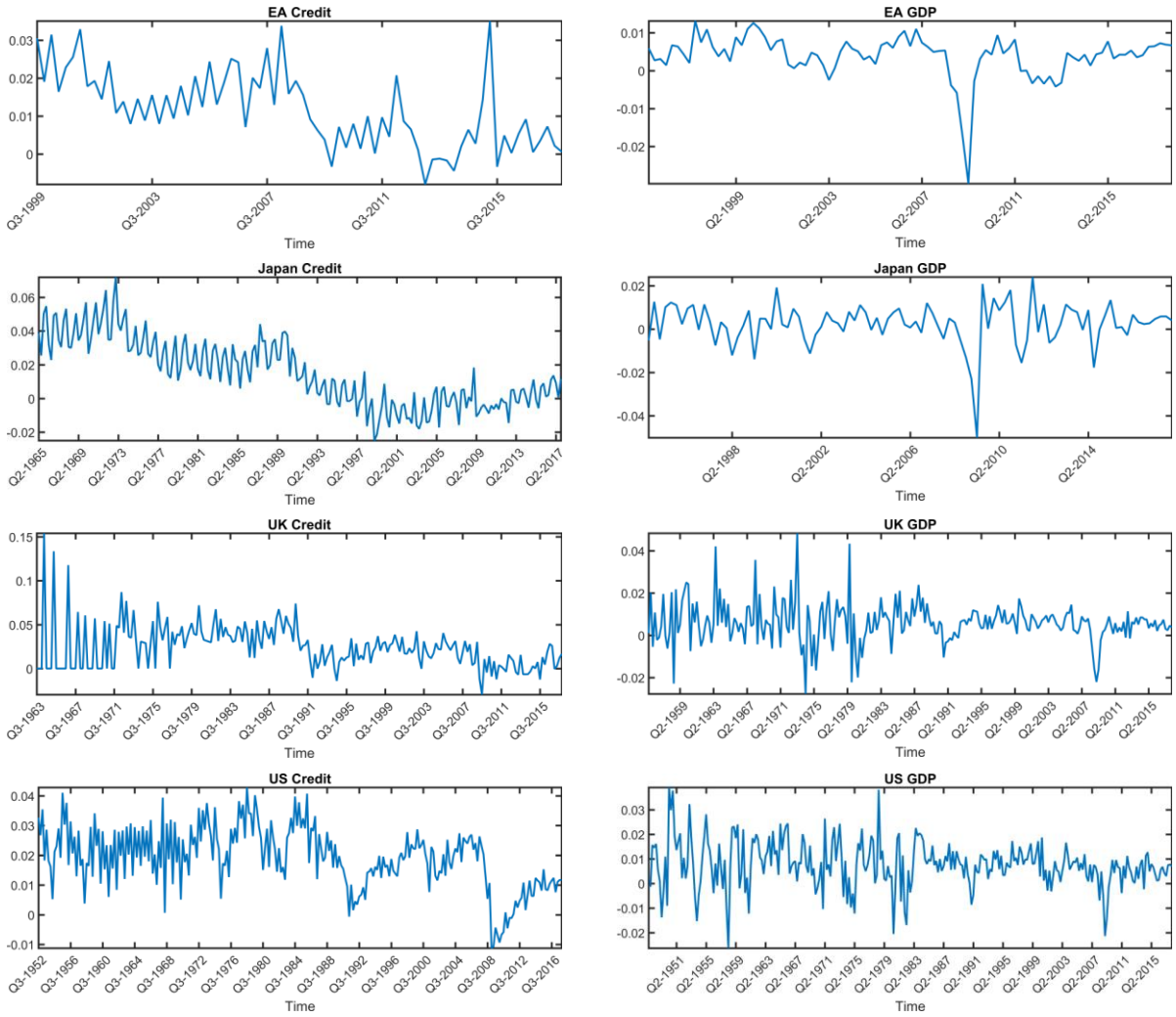


Figure 1 Time series representation, Source: OECD [14], BIS [3]

3 Methods

The relation between two time series can be measured via several approaches in the time-frequency domain. A popular method in economic applications is wavelet cross-spectrum or coherence (Croux *et al.* [5]). For a time series s(t) we can model wavelet spectra (Aloui et al. [1]) as

$$W_s(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t)\, \psi\left(\frac{t-b}{a}\right) dt, \quad a > 0, b \in R \tag{1}$$

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. Assume, that we have two time series $x(t)$ and $y(t)$, both independent Gaussian white noise (GWN), with variances σ_x^2 and σ_y^2. Thus, their wavelet cross-spectrum (WCS) is define as

$$C_{xy}(a,b) = W_x(a,b) W_y^*(a,b), \tag{2}$$

where W_x(a,b) and W_y(a,b) are the coefficients of wavelet transform corresponding to the time series $x(t)$ and $y(t)$. The symbol * denotes complex conjugation. Since the WCS is complex, such in our case, when the complex Morlet wavelet are used, we can define wavelet power cross-spectrum (PWCS), as

$$\left|C_{xy}(a,b)\right|^2 = |W_x(a,b)|^2 \left|W_y^*(a,b)\right|^2, \quad (3)$$

where $|W_x\ (a,b)|^2$ and $|W_y\ (a,b)|^2$ are wavelet power spectra. The significance testing of PWCS can be done according to Ge [11][12] (denoted as Ge08) and Torrence and Compo [19]. The testing statistic takes the form of

$$\left|C_{xy}(a,b)\right|^2 / \left(\sigma_x^2 \sigma_y^2\right) \sim 0.25 W_2, \quad (4)$$

where W_2 denotes probability distribution with probability density function $f\ (z)$=0.5 $K_0(z^{1/2})$. Here z is the random variable, and $K_0(z)$ is the modified Bessel function of order zero when the complex wavelets are used [11][12][19][20].

4 Results

With respect to the aim of the paper, we proceed with cross-spectral analysis. We set the scales in correspondence with the range of 1 year to 10 years, with 361 individual scales. As the type of mother wavelet, we use Complex Morlet wavelet with centre frequency f_c=1.5. In all figures representing PWCS between GDP and Credit for selected countries (Fig. 2, tab. 1), the x-axis represents time, y-axis represents specific periods (cycles in years) and z-axis represents values of spectrogram. The figures show a two-dimensional projection of three-dimensional charts. The intensity of each contour represents the relative importance of the different periodicities and time. The yellow curve in all figures indicates a significant area identified by Ge08.

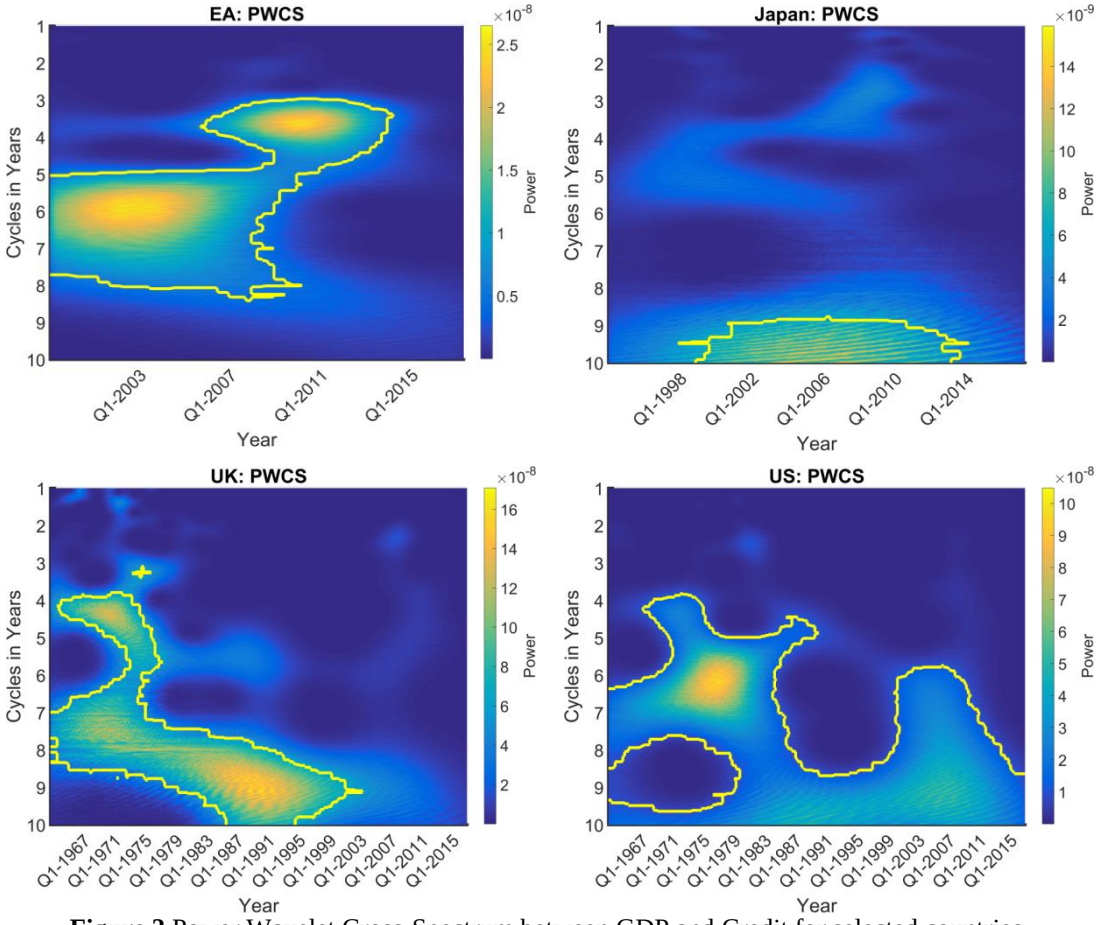


Figure 2 Power Wavelet Cross-Spectrum between GDP and Credit for selected countries.

In all countries, the PWCS with its significance tests show co-movement areas bounded by the time and the frequency (Fig. 2). We can see, that the results are different across countries. There is only one similarity, missing co-movement in cycles lower than 3 years, which represent short cyclical fluctuations and quick reactions. The case of EA shows two basic areas of co-movement, before crisis in 1999/Q-2008/Q4 in long

business cycles (BC) and during and after the crisis in 2006/Q3-2013/Q4 in short and medium business cycles. Both areas overlap the crisis time (2007 and 2008). Situation for Japan is quite different. Here, we can see only long-run cycles over 8 years in 1999/Q4-2013/Q2 with the highest amplitude around the years of the crisis; the standard testing according to Ge08 does not reveal any shorter cyclical co-movements. In the case of the UK, we can find long BC in the period 1964/Q3-1978/Q1 and after that a co-movement in long-run cycles in 1978/Q1-1999/Q2 (cycles of duration over 8 years). The cycles with the highest amplitude can be found around 1991 (9 years long cycles) when the period after the transformation of the UK economy (1980-1996) began under the Margaret Thatcher economic policy measures. The crisis of 2007 and 2008 is not captured in our co-movement analysis by the standard Ge08 test, which is calculated with respect to the given time range. In case of the last country – the US – we can see a significant co-movement during all time range, i.e. in long BC in 1963/Q2-1985/Q4, 2001/Q2-2011/Q4 and in long-run cycles in 1963/Q2-2017/Q3. We can see the highest amplitude around 1976 in long BC. The era of the late 1970s, the US economy suffered from energy crisis, slow economic growth, high unemployment and very high inflation coupled with high interest rates. The financial crisis of 2007 and 2008 is also captured in our analysis but the amplitude of cycles in this period is lower and is moving to the long-run frequencies.

	Short-run cycles (<1.5 years)	**Short and medium business cycles (1.5-4years)**	**Long business cycles (4-8 years)**	**Long-run cycles (>8 years)**
EA	--	2006/Q3-2013/Q4	1999/Q-2008/Q4	--
Japan	--	--	--	1999/Q4-2013/Q2
UK	--	--	1964/Q3-1978/Q1	1978/Q1-1999/Q2
US	--	--	1963/Q2-1985/Q4 2001/Q2-2011/Q4	1963/Q2-2017/Q3

Table 1 Significant area of PWCS over intervals of cycles

Note: statistically significant at 5% according to Ge08 with respect to the available time range. BC means business.

As we already mentioned, Rünstler and Vlekke [16] use multivariate unobserved components models to analyse cyclical character of GDP. They found large and persistent cycles (approx. 15 years) in real credit volumes and real house prices which are highly correlated with a medium BC. And, that financial cycles are larger and longer for countries with high rates of private home ownership. As they state, it suggests possible gains from the coordination of monetary and macro-prudential policies. Igan et al. [13] focus on macro-financial linkage and the cross-country diversity of cyclical co-movements in different way. They use dynamic generalized factor model and spectral techniques in frequency domain. They found that house price cycles are found to lead credit and real activity over the long term, while in the short to medium term the relationship varies across countries. Comparing our results with these papers we can state that we confirmed such findings via different methodology approach. We proved existence of long BC and long-term cycles based on time-frequency co-movement analysis of GDP and financial time series.

5 Conclusion

The paper was focused on the interrelationship between the financial and the macroeconomic sector, i.e. the financial and business cycles, of the selected countries with the regime of quantitative easing (i.e. the euro area, Japan, the United Kingdom and the United States) to find whether these economies face the same co-movement of financial and business cycles. We used credit volumes as a key financial cycle variable (as such, we focus on credit cycles) to examine how the financial cycle differs from the business cycle of a selected country in a set of countries which adopted the regime of quantitative easing.

Our results confirm that there are differences concerning the co-movement of financial and business cycles in analysed countries. We identify the only one similar feature, which is missing co-movement lower than 3 years. The widest range of co-movement in the cross-section of time and frequency can be seen in case of the US economy with large areas of co-movement in long business cycles (4-8 years) and long-run cycles (higher than 8 years) throughout the analysed time range. The situation is slightly similar in the UK economy, but the area of co-movement disappeared after 1999. Euro are faces the cyclical co-movement mainly in short and medium business cycles (1.5-4 years) just before, during and after the crisis and in long business cycles (4-8 years) in the

period before the financial crises, i.e. it is clear that the financial crises had an impact on the character of the euro area cyclical co-movement in the sense that the it became shorter. Japan is the only country with the area of co-movement in long-run cycles. To sum up, euro area is the only economic region where we did not identify the existence of co-movement of financial and business cycles longer than 8 years. However, this finding could be explained by the fact that the data available for the euro area are very limited and as such we cannot measure the co-movement before the euro adoption (i.e. before 1999).

We confirm the results of Rünstler and Vlekke [16] who prove the existence of long financial cycles and the correlation with the medium-term business cycles (longer than common 8-32 quarters used for the definition of business cycle) and identify country differences. National institutional differences influencing the character of medium business cycles is also confirmed by Igan et al. [13].

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] Aloui, Ch., Hkiri, B. & Nguyen D.K. (2016). Real growth co-movement and business cycle synchronization in the GCV countries: Evidence from time-frequency analysis. *Economic Modelling, 52*, 322–331.
[2] Aguiar-Conraria, L., Brinca, P., Gudjonsson, H. & Soares, J. (2015). Optimal currency area and business cycle synchronization across U.S. states. *MPRA working paper, number 62125*. Retrieved from: https://mpra.ub.uni-muenchen.de/62125/1/MPRA_paper_62125.pdf
[3] Bank for International Settlements: *Credit to the non-financial sector* [online database]. BIS (2018) [cit. 2018-04-18]. Retrieved from: https://www.bis.org/statistics/totcredit.htm.
[4] Berdiev, A. N. & Chang, Ch-P. (2015). Business cycle synchronization in Asia-Pacific: New evidence from wavelet analysis. *Journal of Asia Economics, 37*, 20–33.
[5] Claessens, S., Kose, M. A. & Terrones, M. E. (2012). How do business and financial cycles interact? *Journal of International Economics, 87*, 178–190.
[6] Croux, Ch., Forni, M. & Reichlin, L. (2001). A measure of comovement for economic variables: Theory and empirics. *The Review of Economics and Statistics, 83*, 232–241.
[7] Drehmann, M., Borio, C. & Tsatsaronis, K. (2012). Characterising the financial cycle: don't lose sight of the medium term! *BIS Working Paper, 380*.
[8] Fidrmuc, J., Korhonen, I. & Poměnková, J. (2014). Wavelet spectrum analysis of business cycles of China and G7 countries. *Applied Economics Letters, 21*, 1309–1313.
[9] Ftiti, Z., Tiwari, A. & Belanés, A. (2014). Tests of financial market contagion: Evolutionary cospectral analysis v.s. wavelet analysis. *Computational Economics, 46*, 575–611.
[10] Galati, G., Hindrayanto, I., Koopman S. J. & Vlekke, M. (2016). Measuring financial cycles in a model-based analysis: Empirical evidence for the United States and the euro area. *Economic Letters, 145*, 83–87.
[11] Ge, Z. (2008). Significance tests for the wavelet cross spectrum and wavelet linear coherence. *Annales Geophysicae, 26*, 3819–3829.
[12] Ge, Z. (2013). Significance tests for the wavelet power and the wavelet power spectrum. *Annales Geophysicae, 31***,** 315–315.
[13] Igan, D. Kabundi, A., De Simone, F. N., Pinheiro, M. & Tamirisa, N. (2011). Housing, credit, and real activity cycles: Characteristics and comovement. *Journal of Housing Economics, 20*, 210–231.
[14] Organisation of Economic Co-operation and Development: *OECD.Stat: Quarterly National Accounts* [online database]. OECD (2018) [cit. 2018-04-18]. Retrieved from: http://stats.oecd.org/#.
[15] Rua, A. (2012). Money growth and inflation in the Euro area: A time-frequency view. *Oxford Bulletin of Economics and Statistics, 74*, 0305–9049.
[16] Rünstler, G. & Vlekke, M. (2016). Business, housing, and credit cycles. *Journal of Applied Econometrics, 33***,** 212–226.
[17] Schüler, Y. S., Hiebert, P. P. & Peltonen, T. A. (2015). Characterising the financial cycle: a multivariate and time-varying approach. *ECB Working Paper, 1846*.
[18] Tiwari, A. K., Mutascu, M. I. & Abulescu, C. T. (2016). Continuous wavelet transform and rolling correlation of European stock markets. *International Review of Economics and Finance, 42*, 237–256
[19] Torrence, Ch. & Compo, G. P. (1998) A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society, 79*, 61–78.
[20] Wells, W. T., Anderson, R. L. & Cell, J. W. (1962). The distribution of the product of two central or non-central chi-square varieties. *Annals of Mathematical Statistics, 33*, 1016–1020.

Algebraic and graphic comparison of crisp, interval and fuzzy approach to solving linear programming problems

Andrea Kubišová[1]

Abstract. In conventional teaching LP problems, all coefficients of a mathematic model are usually given exactly and an exact optimum can therefore be found. In practice, however, the limit values are often expressed vaguely and it is necessary to use relaxed capacity limits and a more appropriate concept of solving LP problems, especially the fuzzy method.
The article presents a didactic proposal of how to familiarize business students with the idea of interval and fuzzy approaches. It suggests introducing the approaches to the students through illustration and comparison between algebraic and graphic solutions by means of oblique and orthographic projection.
For that reason, a common example with just two structure variables is solved and depicted. The paper will use triangular fuzzy numbers to facilitate the explanation. Our own formulas for interval optimum are derived and Bellman and Zadeh´s min-max operator for fuzzy optimum is used. The graphic solutions are depicted and compared.

Keywords: interval linear programming, fuzzy linear programming, orthographic projection, oblique projection.

JEL Classification: C44
AMS Classification: 90C70

1 Introduction

In Operation Research course for finance students, methods for solving common LP problems are usually lectured. All the coefficients of a scholar mathematic model are commonly given exactly and an exact optimum can therefore be found. From didactic point of view, it is advantageous to construct the graphic solution parallel to introducing the one- and two-phase primary simplex methods and to compare the steps in equivalent matrices with geometry objects in the picture.

In practice, this usually does not refer to real economic problems; it is used as the simplest way of its modeling especially for scholar use. During the simplex method explanation, it is usual and useful to compare the progression in the simplex table with the points in the graph.

However, especially the capacity limit values are not expressed exactly and for real situations modeling it is necessary to use relaxed coefficients or constraint limits, which requires an appropriate concept of solving LP problems. Our aim is to make students familiar especially with the idea of fuzzy method approach, using the interval method as an intermediate stage. Both can be illustrated with the geometric solution again during explanation.

Because of dealing with only graphically solvable problems, only LP problems with two structural variables are considered in all this paper.

2 Vague capacities of LP problems

In this article, only relaxation for constraint limits is considered. The crisp, interval and fuzzy number will be used in the next three paragraphs. The number is defined, links to the author´s former articles where formulas for algebraic solution are derived are made and the method for illustrative graphic solution is suggested.

2.1 Crisp limits of capacities

When teaching LP problems solution conventionally, we construct mathematic models including constraints with exact limits on the right-hand side labelled b_1 and b_2

$$\begin{aligned} a_{11}x_1 \quad + a_{12}x_2 \quad &\leq \quad b_1 \\ a_{21}x_1 \quad + a_{22}x_2 \quad &\leq \quad b_{2.} \end{aligned} \tag{1}$$

[1] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, Jihlava, andrea.kubisova@vspj.cz

Thus, the exact nonnegative optimum of objective function $z_{opt} = c_1x_1 + c_2x_2$ is always gained.

Algebraic solution

The classic theory of solving LP problems using simplex method is introduced in detail in Bazaraa (2011). The following formulae for expressing the crisp optimal solution

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \text{ a } x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \tag{2}$$

are derived and after their substitution to

$$z_{opt} = c_1x_1 + c_2x_2, \tag{3}$$

an appropriate objective function value is gained as it is shown in Kubišová (2015c). The symbols used in (2) and (3) correspond to Figure 1.

Graphic solution

A Gaussian coordinates system with only two dimensions will for depicting all those crisp values graphically. Values of the structure variables correspond to the values on the axes. Each crisp solution is depicted as a point. Each crisp constraint border is depicted as a line, each constraint as a half-plane.

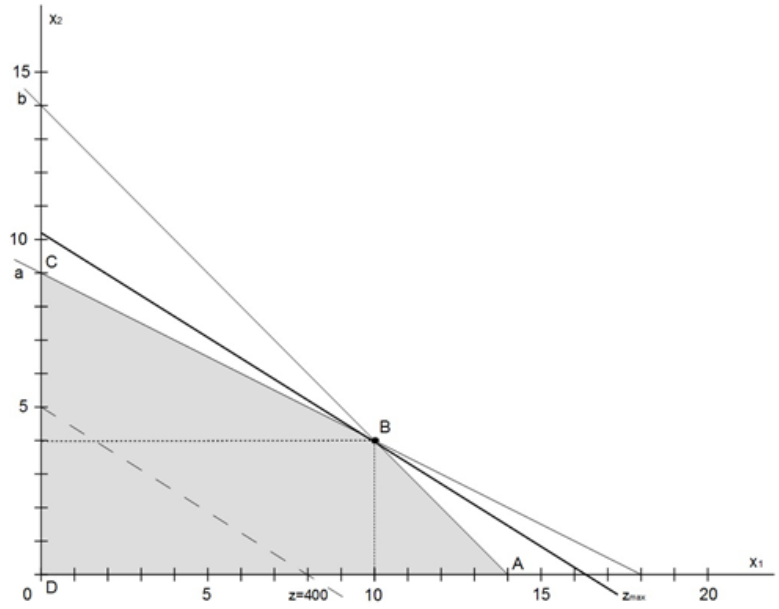


Figure 1 Crisp LP problem – optimal solution

The feasible region is a convex polygon, the isoquants of the same value of objective function make a system of parallel lines, the most distant one from its zero level still having nonempty intersection with the feasible region labelled z_{max}, determines the optimal solution point position, labelled as *B* in Figure 1.

2.2 Interval limits of capacities

Real LP problems where the limit values are not precisely known can be simplified and modeled with relaxed interval capacity limits. The interval approach replaces crisp capacities with interval numbers on the right-hand sides marked as B_1, B_2, as introduced in Moor (2009)

$$\begin{aligned} a_{11}x_1 \quad + a_{12}x_2 \quad &\leq \quad B_1 \\ a_{21}x_1 \quad + a_{22}x_2 \quad &\leq \quad B_2\,. \end{aligned} \tag{4}$$

Then a nonnegative interval optimum of the objection function $z_{opt} = c_1x_1 + c_2x_2$ is gained according to the definitions.

Although this interval approach is not used in practice, because it still does not describe the real situation properly, it may help the students understand the point of using and operating the fuzzy capacities. As an intermediate step between crisp and fuzzy approach, each interval number can be understood as a special fuzzy number where all its values have the same level of the membership function.

Algebraic solution

The following formulae for expressing interval solution

$$X_1 = \frac{\begin{vmatrix} B_1 & a_{12} \\ B_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}},\ X_2 = \frac{\begin{vmatrix} a_{11} & B_1 \\ a_{21} & B_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \tag{5}$$

and the appropriate objective function value

$$Z_{\max} = \frac{1}{|\boldsymbol{A}|}\left[c_1 \begin{vmatrix} b_1'' & a_{12} \\ b_2' & a_{22} \end{vmatrix} + c_2 \begin{vmatrix} a_{11} & b_1' \\ a_{21} & b_2'' \end{vmatrix}; c_1 \begin{vmatrix} b_1' & a_{12} \\ b_2'' & a_{22} \end{vmatrix} + c_2 \begin{vmatrix} a_{11} & b_1'' \\ a_{21} & b_2' \end{vmatrix} \right] \tag{6}$$

are derived in Kubišová (2015c). The symbols used in (5) and (6) correspond to Figure 2.

Graphic solution

Gaussian coordinates system with only two dimensions will be sufficient for depicting interval values graphically again, values of the structure variables correspond to the values on the axes. Each interval number is depicted as a line segment.

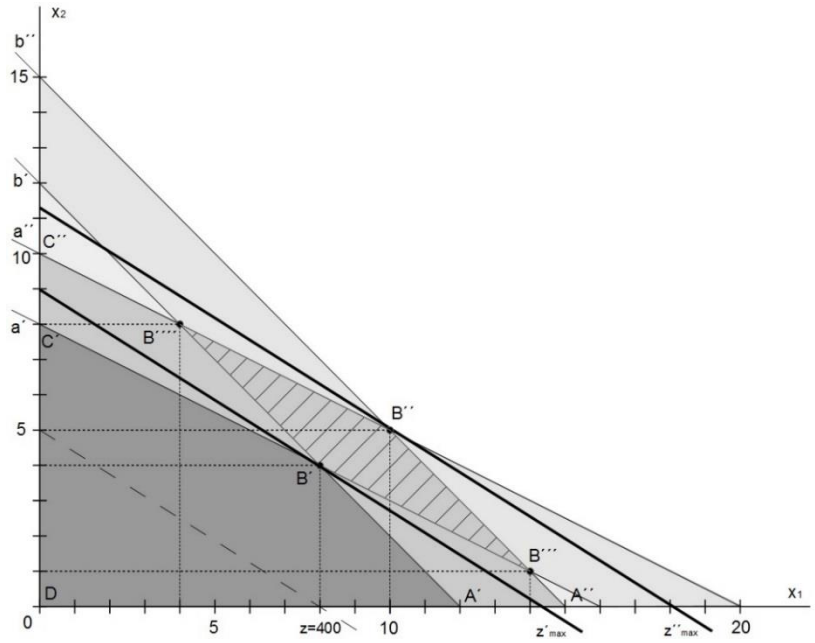


Figure 2 Interval LP problem

Interval solution is depicted as a set of points with both coordinates given as an interval, which always creates a parallelogram. The interval optimal solution is depicted as a hatched parallelogram B´B´´´B´´B´´´´ in Figure 2. Each interval constraint border is depicted as a strip, each constraint as a half-plane with this border.

2.3 Fuzzy limits of capacities

Nowadays especially the fuzzy approach is used as the most appropriate concept for expressing the vague and not exactly given numbers, the same applying to the LP problem constraint limits.

The fuzzy approach replaces crisp capacities with fuzzy numbers on the right-hand sides marked as $\widetilde{B_1}$, $\widetilde{B_2}$, as introduced e.g. in Zadeh (1975) or Klir (2005)

$$\begin{array}{lll} a_{11}x_1 + a_{12}x_2 \leq \widetilde{B_1} \\ a_{21}x_1 + a_{22}x_2 \leq \widetilde{B_2}. \end{array} \tag{7}$$

Then a nonnegative interval optimum of the objective function $z_{opt} = c_1x_1 + c_2x_2$ is gained according to the definitions.

Algebraic solution

The fuzzy LP problem can be rewritten as an equivalent crisp LP problem with α as a parameter using **Bellman and Zadeh´s min-max operator** (in [2]) as it is shown in Kubišová (2015b) or **the α-cuts method** can be used if an objective function minimal value requirement is added (Zadeh (1975), Klir (2005)). The following formulae for the membership functions of interval solution only for the x_1 unknown is shown below

$$\mu_{X_1}(x) = \begin{cases} \dfrac{|A|x - \begin{vmatrix} b_1' & a_{12} \\ b_2'' & a_{22} \end{vmatrix}}{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix} - \begin{vmatrix} b_1' & a_{12} \\ b_2'' & a_{22} \end{vmatrix}}, \dfrac{1}{|A|}\begin{vmatrix} b_1' & a_{12} \\ b_2'' & a_{22} \end{vmatrix} \leq x \leq \dfrac{1}{|A|}\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix} \\ \dfrac{\begin{vmatrix} b_1'' & a_{12} \\ b_2' & a_{22} \end{vmatrix} - |A|x}{\begin{vmatrix} b_1'' & a_{12} \\ b_2' & a_{22} \end{vmatrix} - \begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}, \dfrac{1}{|A|}\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix} \leq x \leq \dfrac{1}{|A|}\begin{vmatrix} b_1'' & a_{12} \\ b_2' & a_{22} \end{vmatrix} \end{cases}. \tag{8}$$

The formulae for its objective function value requires more space, mentioned formulae are derived in Kubišová (2015c). The symbols used in (8) correspond to Figure 3.

Graphic solution

Unlike the crisp or interval numbers, whose value of the membership function is always equal to 1 and the two dimensional Gaussian coordinates system is sufficient to depict the relations, the value of membership function of fuzzy numbers is needed to be expressed somehow in the graph.

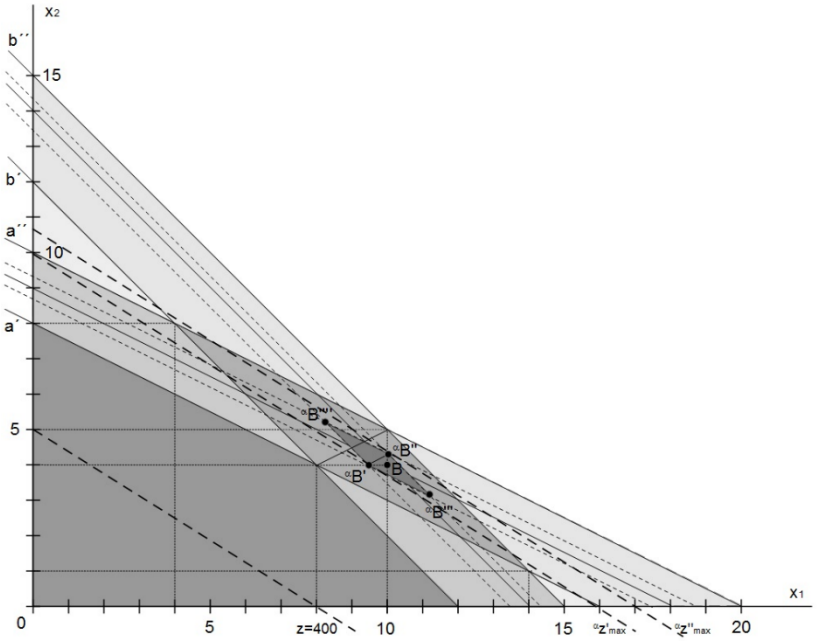


Figure 3 Fuzzy LP problem – optimal solution in the orthographic projection

It requires using descriptive geometry means. This article suggests two geometry methods that are quite simple to draw and the most demonstrative and illustrative: the orthographic projection and the oblique projection.

The orthographic projection enables the students to solve the two-dimensional problem in an undistorted ground plane, the level of the membership function is expressed as the third dimension (the altitude usually written in brackets). Using just triangular fuzzy numbers, where each number is depicted only as a segment with the side points (level 0) and the marked highest point (level 1), is sufficient. The inner points membership function value increases linearly, there is therefore no need to write altitudes.

All fuzzy numbers fitting to the border equations of the linear fuzzy constraint form a set of points representing their union which can be depicted as a strip of a skewed plane containing points at the highest membership function level height. In the graph, the border lines at the zero level are dashed and double-dashed, the points with the top membership function at the level of 1 form a middle line labelled without dashes. All three of them are depicted as parallel lines. The fuzzy feasible region made as an intersection of all fuzzy constraints is depicted as the top surface of a polyhedron with skewed sides representing the vague borders and a height of 1, and the rest of the space at zero height representing a set of the infeasible solutions.

Each fuzzy optimal solution is depicted as a set of points with both coordinates given as a fuzzy number, the α-cuts of this solution form a system of homothetic parallelograms with the center in the *B* point. The situation for the membership value of the optimal solution at least at the level of $\alpha = 0.8$ is highlighted with the dark color and labelled B′B′′B′′B′′′′ in Figure 3.

The oblique projection uses complicated geometry means, but gives a more illustrative view of the situation in 3D view, especially in the two-dimensional problem leading to the usage of Bellman and Zadeh´s min-max operator. The membership function of each number is expressed with the height again, but as a third coordinate μ in this case.

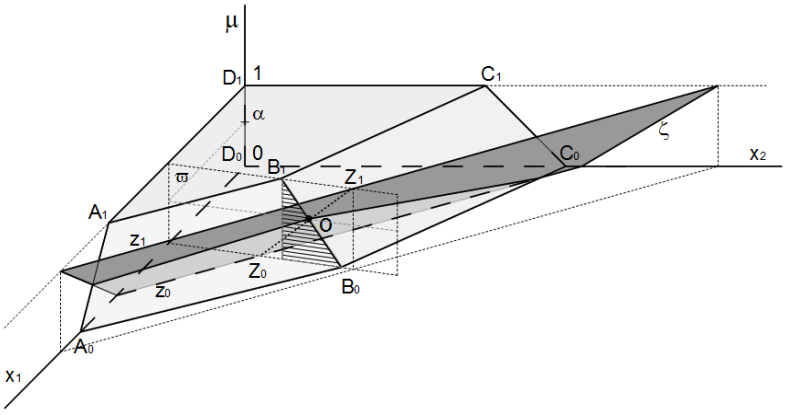


Figure 4 Fuzzy LP problem – optimal solution in the oblique projection

Each fuzzy constraint border is depicted as a top surface of a triangular prism on the strip of the border. The feasible region is a union of all fuzzy constraints, represented by the light colored top surface of a non-zero polyhedron with skewed sides made of three parts $A_0B_0B_1A_1$, $B_0C_0C_1B_1$, C_0D_0, $A_1B_1A_1D_1$ and the corresponding parts of horizontal planes at the height of 0 or 1 in Figure 4.

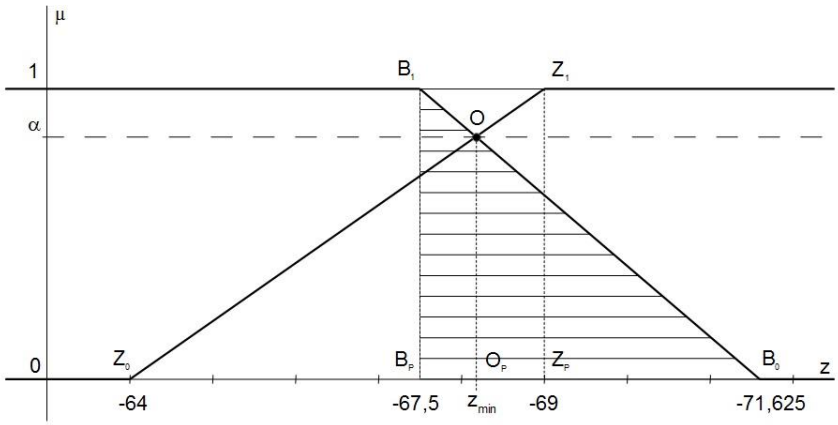


Figure 5 Fuzzy LP problem - optimal solution in a vertical cut

If requirement of a particular level of objective function is added (from unacceptable values through fully acceptable ones), it can be depicted as a skewed strip of plane ξ and horizontal half-planes beyond the acceptance values at the 0 resp. 1 height in Figure 4 or as a line Z_0Z_1 in Figure 5.

The highest point of the intersection line of those two surfaces (the fuzzy feasible region and the objective function with the requirement) determines the optimal solution with the maximal reachable membership function value, in Figures 4 and 5 labelled as *O*, which can be easily found in an undistorted vertical cut plane ω, as depicted in Figure 5.

3 Conclusion

This article suggests how to make the students more familiar with LP problems with fuzzy constraint capacities approach by descriptive geometry means. Using the orthographic projection (the simplest orthogonal one) is recommended for crisp, interval and fuzzy in α-cut form constraint limits, while the oblique projection (a more illustrative one, but still not too difficult to proceed) is more appropriate especially for construction of intersections in a fuzzy LP problem with objective function minimum value requirement added.

Intentionally, particular scholar problems with similar tasks are solved by means of their algebraic as well as graphic solutions to be easily compared. In the cited articles, the author defines the fundamental interval and fuzzy numbers arithmetic operations along with the formulae for the optimal solution, deriving its objective function values.

References

[1] Bazaraa, M. S., Jarvis, J. J. & Sharali, H. D. (2011). *Linear Programming and Network Flows*. New York: John Wiley & sons, Inc.

[2] Gen, M. & Cheng, R. (2000). *Genetic Algorithms and Engineering Optimization*. New York: John Wiley & sons, Inc.

[3] Klir, G. J. & Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. New Delhi: Prentice Hall of India Private limited.

[4] Klir, G. J. & Yuan, B. (2005). *Fuzzy Sets and Fuzzy Logic*. New Delhi: Prentice Hall of India Private limited.

[5] Kolesárová, A. & Kováčová, M. (2002). *Fuzzy množiny a ich aplikácie*. Bratislava: Slovenská technická univerzita.

[6] Kubišová, A. (2015a) Řešení úloh LP s fuzzy kapacitami vlastních omezení metodou α-řezů. *Logos polytechnikos, 6*, 85 – 108.

[7] Kubišová, A. (2015b). Fuzzy limits of capacities in constraints of linear programming problems. *Logos polytechnikos, 6*, 256-274.

[8] Kubišová, A. (2015c). Úloha LP při intervalově zadaných kapacitách vlastních omezení. *Logos polytechnikos, 6*, 68-84.

[9] Lai, Y. J. & Hwang, C. L. (1992). *Fuzzy Mathematical programming: Theory and Applications*. Berlin: Springer.

[10] Moor, R. E., Kearfott, R. G. & Cloud, M. J. (2009). *Introduction to Interval Analysis*. Philadelphia: SIAM.

[11] Navara, M. & Olšák, P. (2007). *Základy fuzzy množin*. Praha: Skriptum ČVUT.

[12] Novák, V. (2000). *Základy fuzzy modelování*. Praha: BEN.

[13] Orlovsky, S.A. (1978). Decision making with fuzzy preference relation. *Fuzzy Sets and Systems, 13*, 155-167.

[14] Ramík, J. & Římánek, J. (1985). Inequality relation between fuzzy numbers and its use in fuzzy optimization. Fuzzy *Sets and Systems, 16*, 123-138.

[15] Tanaka, H. & Asai, K. (1984). Fuzzy linear programming with fuzzy numbers. Fuzzy *Sets and Systems, 13*, 1-10.

[16] Zadeh, Lotfi A., Fu K., Tanaka, K. & Shimura, M. (1975). F*uzzy sets and their applications to cognitive and decision processes.* New York: Academic Press, Inc.

[17] Zimmermann, H. – J. (1978). Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and Systems, 1*, 45-55.

Modeling heavy tail property of financial asset returns with skewed generalized t– distribution

Jaromír Kukal[1], Quang Van Tran[2]

Abstract. The heavy tail property of financial asset returns is well known and it can be modeled by alpha stable distributions or some distribution from the generalized hyperbolic distribution family. However, the first group may often have too heavy tails and the second group exhibits too light ends at times as they are sometime called semi-heavy tail. In this paper, we examine the ability of several distributions from the skewed generalized t-distribution class to capture the fat tail feature of a group of stock market indices. The chosen distributions form the group are skewed generalized t-distribution, and its special cases: generalized t-distribution, skewed t-distribution and t-distribution. In our analysis, first we use maximum likelihood estimation method to estimate parameters of each distribution from data. Then we evaluate their suitability for modeling the distribution of returns of chosen indices using the chi squared goodness of fit test. The test is applied on the whole distribution as well as on some important parts of the distribution. The results of our investigation provide some interesting insights for modeling final asset returns.

Keywords: skewed generalized t-distribution, maximum likelihood estimation, chi squared goodness of fit test, stock market indices

JEL classification: C12, C13

1 Introduction

Heavy tail property of financial asset returns has been addressed in the literature for long time and its deviation from a normal distribution has a principal impact for financial engineering. The alternative distributions chosen for replacing the gaussian distribution are the alpha stable distribution popularized by Nolan [7] or generalized hyperbolic distribution and its special cases [8]. However these distributions have often not fulfilled the expectation for this purpose. Though distributions from the skewed generalized t-distribution class are known for their ability to exhibit fat tail property, they have not attracted so much attention from researchers. To fill this gap in the literature, in this paper we will examine their suitability of skewed generalized t-distribution, and its special cases: generalized t-distribution, skewed t-distribution and t-distribution for capturing heavy tail property of stock market indices returns. For this objective four stock market indices from Central European stock markets are chosen: Budapest stock market index BUX, Vienna stock market index ATX, Warsaw stock market index WIG, and Prague stock market index PX. They are daily series from 2000 to 2018. First we use these series to estimate parameters of these distribution and then the chi squared goodness of fit test is applied to verify whether these distributions can model the heavy tail property present in the chosen data.

2 Skewed Generalized t-Distribution Group

This distribution was first proposed by Theodossiou [10]. It has five parameters: the location $\mu \in \mathbb{R}$, the scale $\sigma > 0$, the skewness $-1 < \lambda < 1$ and the kurtosis $p, q > 0$. Its probability density function is

$$\mathrm{f}(x) = p\left[2v\sigma q^{\frac{1}{p}}\mathrm{B}\left(\frac{1}{p},q\right)\right]^{-1}\left[\frac{|\,x-\mu+m\,|^p}{q(v\sigma)^p(\mathrm{sign}(x-\mu+m)\lambda+1)^p}+1\right]^{-\left(\frac{1}{p}+q\right)}, \tag{1}$$

where

$$m = \frac{2v\sigma\lambda q^{\frac{1}{p}}\mathrm{B}\left(\frac{2}{p},q-\frac{1}{p}\right)}{\mathrm{B}\left(\frac{1}{p},q\right)},$$

[1]Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering, Trojanova 13, Praha 2 - 120 00, jaromir.kukal@fjfi.cvut.cz

[2]Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering, Trojanova 13, Praha 2 - 120 00, tranvqua@fjfi.cvut.cz

$$v = q^{-\frac{1}{p}} \left[\sqrt{(3\lambda^2+1)\frac{\mathrm{B}\left(\frac{3}{p}, q-\frac{2}{p}\right)}{\mathrm{B}\left(\frac{1}{p}, q\right)} - 4\lambda^2 \frac{\mathrm{B}^2\left(\frac{2}{p}, q-\frac{1}{p}\right)}{\mathrm{B}^2\left(\frac{1}{p}, q\right)}} \right]^{-1},$$

and $\mathrm{B}(.,.)$ is the so called beta function. It has mean of μ if $pq > 1$ and variance of σ^2 if $pq > 2$.

2.1 Skewed t-distribution

It is a subclass of skewed generalized t-distribution discovered by MacDonald and Newey [6] for $p = 2$. Hence it has only four parameters. Its PDF is

$$\mathrm{f}(x) = \frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\kappa\sigma\sqrt{\pi\nu}\Gamma(\nu)} \left(\frac{(x-\mu+m)^2}{\nu[\kappa\sigma(\beta \operatorname{sign}(x-\mu+m)+1)]^2} + 1 \right)^{-\left(\nu+\frac{1}{2}\right)}, \tag{2}$$

where

$$m = \frac{2\kappa\sigma\sqrt{\nu}\Gamma\left(\nu-\frac{1}{2}\right)}{\sqrt{\pi}\Gamma\left(\nu+\frac{1}{2}\right)} \tag{3}$$

and

$$\kappa = \left[\nu \left(\left(3\beta^2+1\right)\frac{1}{2\nu-2} - \frac{4\nu^2}{\pi} \left(\frac{\Gamma\left(\nu-\frac{1}{2}\right)}{\Gamma(\nu)} \right)^2 \right) \right]^{-\frac{1}{2}}. \tag{4}$$

The analytical form for the corresponding CDF is unknown.

2.2 Generalized Student t-distribution

when $\lambda = 0$, we obtain the generalized Student t-distribution, whose PDF is:

$$\mathrm{f}(x) = \frac{p}{2\nu\sigma q^{1/p}\mathrm{B}\left(\frac{1}{p}, q\right)} \left[\frac{|x-\mu|^p}{q(\nu\sigma)^p} + 1 \right]^{-\left(\frac{1}{p}+q\right)}, \tag{5}$$

where

$$v = q^{-1/p} \sqrt{\frac{\mathrm{B}\left(\frac{1}{p}, q\right)}{\mathrm{B}\left(\frac{3}{p}, q-\frac{2}{p}\right)}}. \tag{6}$$

The CDF of this distribution is also unknown.

2.3 Student t-distribution

The Student t distribution has three parameters: location $\mu \in \mathbb{R}$, scale $\sigma > 0$ and the so called the number of degrees of freedom $\nu > 0$. The third parameter ν just defines the tail property of this distribution. The higher value ν attains, the lower value the kurtosis of this distribution is. Its PDF is

$$\mathrm{f}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(x-\mu)^2}{\nu\sigma^2} \right)^{-\frac{\nu+1}{2}} \tag{7}$$

for where $x \in \mathbb{R}$ and the corresponding CDF is

$$\mathrm{F}(x) = \frac{1}{2} + \frac{x-\mu}{\sigma}\Gamma\left(\frac{\nu+1}{2}\right) \frac{{}_2\mathrm{F}_1\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{(x-\mu)^2}{\nu\sigma^2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \tag{8}$$

where Γ is the so called gamma function and ${}_2\mathrm{F}_1$ is the hypergeometric function.

3 Empirical Analysis and Results

In order to investigate the suitability of distributions from the skewed generalized t-distribution group, we use four stock market indices from Central Europe countries of Hungary (BUX), Austria (ATX), Poland (WIG) and the Czech Republic (PX). All four original stock market index series from the beginning of 2000 to February of

	Original series				Log-returns series			
	BUX	ATX	WIG	PX	BUX	ATX	WIG	PX
mean	1.84e4	2431.2	3.82e4	985.1	3.23e-4	2.31e-4	2.63e-4	1.82e-4
median	1.87e4	2388.9e3	4.13e4	979.4	4.99e-4	7.38e-4	4.72e-4	6.05e-4
mode	1.83e4	1127.8	1.37e4	407.7	-0.126	-0.103	-8.47e-2	-1.27e-2
min	5671.0	1003.7	1.16e4	320.1	-0.126	-0.103	-8.47e-2	-0.162
max	4.15e4	4981.9	6.76e4	1936.1	0.132	0.120	6.08e-2	0.124
std	7829.1	970.0	1.53e4	361.5	1.51e-2	1.43e-2	1.25e-2	1.38e-2
skew	0.385	0.588	-0.210	0.328	-0.100	-0.334	-0.381	-0.478
kurt	2.94	2.81	1.86	2.78	9.29	10.1	6.48	16.0
obs	4735	4735	4735	4554	4538	4493	4547	4538

Table 1 Descriptive statistics of original and logarithmic transformed series

2018 are converted into log-return series. The descriptive statistics of both original series and log-return series are shown in Table 1 below. The logarithmic return series are first used to estimate the parameters of the investigated distributions with maximum likelihood estimation technique. All calculations are performed with Matlab and the computational procedure are very highly extensive. While the estimation procedure for t-distribution and skewed t-distribution are standard routines, the estimation of parameters of generalized t-distribution and skewed generalized t-distribution are challenging as the procedure contains some non-linear constraints. The estimation results for each investigated distribution are reported in Tables 2-5. Besides the estimates themselves, their corresponding asymptotic standard errors are also calculated. They are shown on the right-hand side of the estimates in the tables.

We use the estimation results to compute the theoretical probability density of the examined distributions for

	BUX		ATX		WIG		PX	
	coeff	S.E.	coeff	S.E.	coeff	S.E.	coeff	S.E.
μ	3.99e-4	3.71e-2	8.18e-4	1.48e-4	4.86e-4	1.38e-4	5.33e-4	1.56e-4
σ	1.08e-2	2.16e-4	9.08e-3	1.26e-4	8.91e-3	1.14e-4	8.87e-3	1.77e-4
ν	4.08	0.241	3.12	2.46e-2	3.77	1.71e-2	3.37	3.96e-2

Table 2 Estimation results for Student t-distribution

	BUX		ATX		WIG		PX	
	coeff	S.E.	coeff	S.E.	coeff	S.E.	coeff	S.E.
μ	3.59e-4	3.41e-3	5.13e-4	4.42e-3	4.51e-4	4.62e-2	4.03e-4	1.75e-4
σ	1.51e-2	7.27e-4	1.51e-2	2.93e-3	1.30e-2	4.05e-4	1.39e-2	3.13e-4
ν	1.83	3.45e-2	1.59	4.96e-3	1.89	6.91e-2	1.70	5.46e-2
λ	-9.34e-3	2.20e-2	-8.81e-2	4.32e-3	-3.22e-2	1.91e-2	-5.69e-2	2.05e-2

Table 3 Estimation results for skewed t-distribution

	BUX		ATX		WIG		PX	
	coeff	S.E.	coeff	S.E.	coeff	S.E.	coeff	S.E.
μ	4.09e-4	1.95e-4	8.18e-4	2.99e-4	4.88e-4	2.93e-4	5.34e-4	8.44e-3
σ	1.51e-2	3.23e-4	1.50e-2	8.47e-4	1.26e-2	3.69e-4	1.38e-2	1.88e-4
p	1.89	6.03e-2	1.94	2.91e-2	1.46	6.83e-2	1.96	1.47e-2
q	2.34	0.276	1.66	0.168	4.62	0.398	1.76	2.30e-2

Table 4 Estimation results for generalized t-distribution

each log-return series. These theoretical probabilities are then compared to their empirical density. The results are shown in Figures 1-4. On the left panels whole distributions are displayed and on the right panels the left tail of distribution are zoomed in to show the thickness of the tail of each distribution.

	BUX		ATX		WIG		PX	
	coeff	S.E.	coeff	S.E.	coeff	S.E.	coeff	S.E.
μ	3.18e-4	2.13e-4	2.66e-4	4.80e-3	3.03e-4	1.59e-4	2.17e-4	2.60e-3
σ	1.50e-2	3.57e-4	1.49e-2	6.84e-4	1.26e-2	1.98e-4	1.38e-2	2.73e-4
λ	-1.67e-2	2.09e-2	-8.81e-2	5.54e-3	-3.11e-2	1.43e-2	-5.66e-2	1.94e-2
p	1.88	2.33e-2	1.91	4.92e-2	1.46	4.44e-2	1.95	7.00e-2
q	2.36	6.86e-2	1.74	7.83e-2	4.69	2.68e-1	1.78	6.18e-2

Table 5 Estimation results for skewed generalized t-distribution

The estimation results show that ν, i.e. the number of degrees of freedom of t-distribution, of returns of indices from emerging markets (Hungary, Poland and the Czech Republic) are higher than the one of Austria. This indicates that returns of stock market indices from emerging markets tend to have thinner tails than from a traditional market. This is also confirmed in the case of skewed t-distribution, generalized t-distribution, and skewed generalized t-distribution, where the product pq of returns of three emerging markets is also higher than the one of the Austrian market.

Regarding the skew parameter λ, the estimates are negative and very close to zero in all cases indicating that the distributions of returns of all investigated stock market indices are only slightly skewed to the right. The estimated values of parameter p are close to 2 except the one for index WIG returns (1.46). These results imply that n general all special cases do not substantially deviate from the general case. This finding can be observed in Figures 1-4 where all distributions are very close to each other. As the distributions are skewed, they tend to have slightly thicker tails than the unskewed ones. We also calculate the values of log-likelihood function for each distribution and the corresponding stock market index. The results are reported in Table 6. The results show that the inclusion of additional parameters does not substantially increase the goodness of fit of the analyzed distribution for the selected data except the case of index ATX, when the impact of the skewed parameter is evident [3].

Distribution	BUX	ATX	WIG	PX
Student t	12937.54	13220.62	13740.47	13568.06
Skewed Student t	12932.52	13230.24	13741.81	13572.29
Gen. Student t	12937.36	13220.69	13748.96	13568.01
Skewed Gen.Student t	12938.24	13230.40	13750.50	13572.26

Table 6 The values of log-likelihood function for each distribution

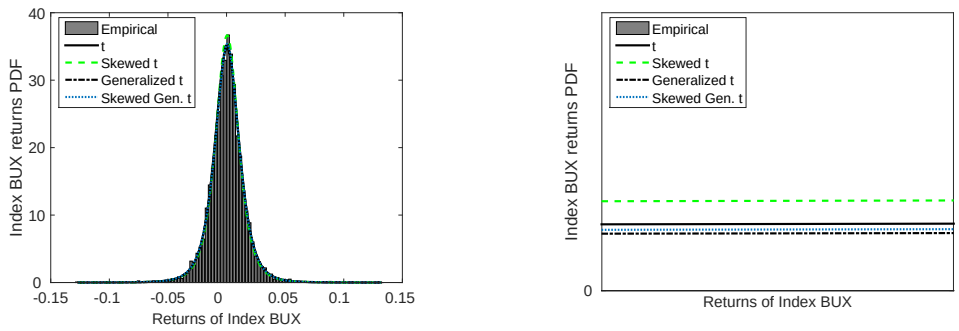


Figure 1 Distribution comparison for returns of index BUX

Besides visual inspection (the chi squared by eye test), we also analyze the suitability of each candidate distribution for modeling returns of stock market indices by a statistic hypothesis testing procedure. There are several possibilities how to test it and the chi squared goodness of fit test is chosen for this purpose as it takes into account the number of parameters of a candidate distribution. The essence of this test is as follows. This test tests the null hypothesis whether data comes from a certain distribution. The measure of goodness of fit, which is also the test

[3] These values can be used to test various restrictions of parameters with likelihood ratio test.

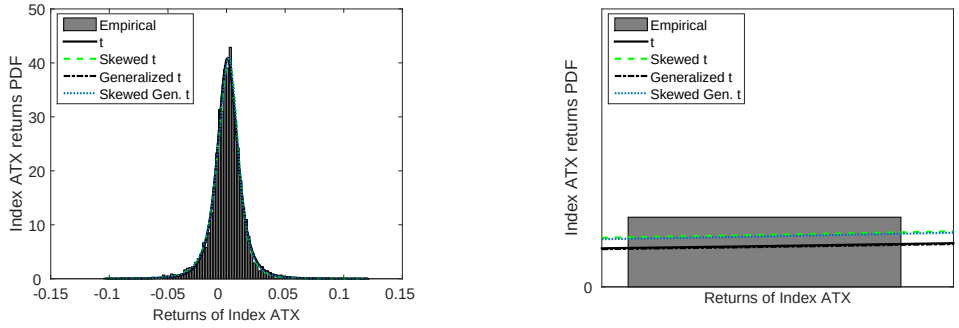


Figure 2 Distribution comparison for returns of index ATX

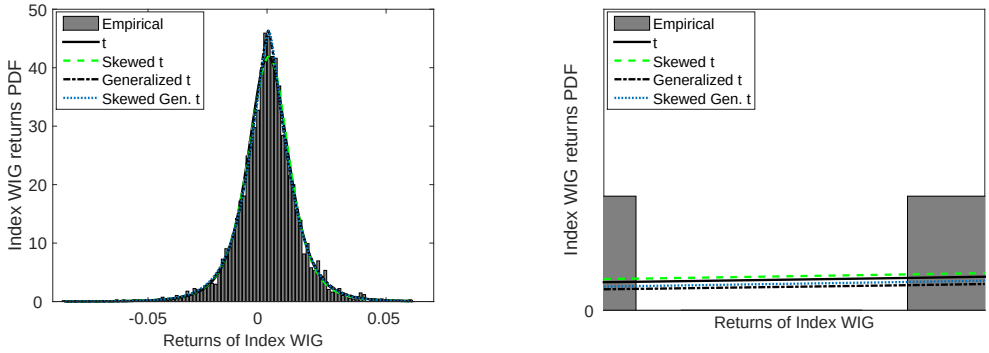


Figure 3 Distribution comparison for returns of index WIG

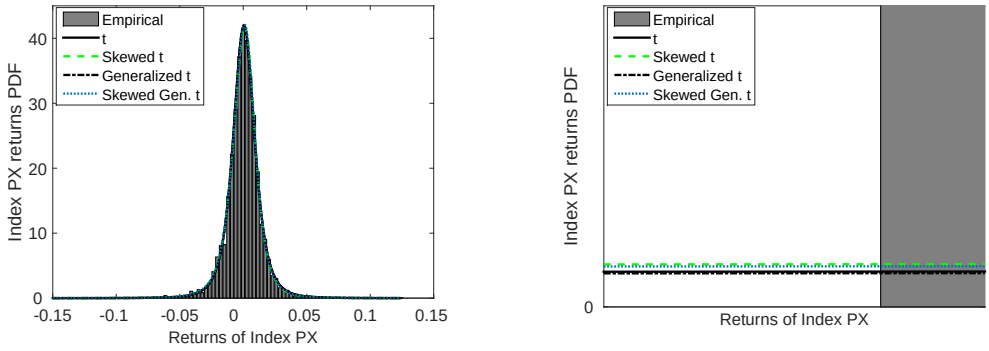


Figure 4 Distribution comparison for returns of index PX

statistic, compares the observed frequencies with the expected ones by summing up their differences as follows

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}, \tag{9}$$

where

- O_i the observed frequency for the i-th bin,
- E_i the expected frequency for this bin,

where E_i is computed as follows

$$E_i = N\left(\mathrm{F}(ub_i) - \mathrm{F}(lb_i)\right), \tag{10}$$

where N is the total number of observations, F is the CDF of the hypothesized distribution, lb_i and ub_i are the lower bound and upper bound of the i-th bin respectively. The test statistic under the null hypothesis has χ^2 with $n-k-1$ degrees of freedom, where n is the number of bins and k is the number of parameters of the tested distribution (see Snedecor and Cochrane, [9]).

The testing is proceeded as follows. Interval [0,1] into forty subintervals of length 0.025. For t-distribution the inverse CDF is used to compute the boundary points for each bin in terms of returns. For the other distribution,

as their CDF are unknown, they are determined numerically. Then the boundary points for each bin are compute with interpolation. After that the number of observed frequencies is counted and the test statistic is calculated according to 9. The results of the chi squared goodness of fit test for all distribution are displayed into Table 7 with the corresponding p-value.

Distribution	BUX		ATX		WIG		PX	
	stat	pval	stat	pval.	stat	pval	stat	pval
t	35.92	0.4721	69.89	6.03e-4	36.22	0.4586	35.02	0.5152
Skewed t	38.68	0.3071	46.62	0.0950	40.40	0.2438	42.23	0.1859
Gen t	36.51	0.3985	67.00	9.02e-4	39.88	0.2622	35.11	0.4643
Sk. Gen. t	37.02	0.3313	46.12	0.0802	39.70	0.2309	45.28	0.0935

Table 7 Chi squared goodness of fit test results

The results clearly indicate that for indices BUX, WIG, and PX the the skewed generalized t-distribution or anyone of its special cases can well capture the heavy tail property of these indices returns. According to the test statistic, in this case the best fit is obtained when the Student t-distribution is applied. It is a very favorable result for financial engineering as for t-distribution, all its PDF, CDF and ICDF are known. For index ATX unskewed t-distribution and generalized t-distribution are not suitable for modeling its returns. The skewed t-distribution or skewed generalized t-distribution provide a much better fit for returns of index ATX. However, the fit is on the border line of acceptability and we may still have to find a more suitable alternative for data of index ATX type.

4 Conclusion

We have investigated the ability of the skewed generalized t-distribution and its three special cases its special cases: generalized t-distribution, skewed t-distribution and t-distribution to model the heavy tail property of financial asset returns. For this purpose four stock market indices BUX, ATX, WIG, and PX from Central European stock markets are chosen. They are daily series from 2000 to 2018. First, the parameters of these distributions have been estimated from data using maximum likelihood estimation method and then the chi squared goodness of fit test is applied to examine their suitability for the chosen data. The results we obtained are quite encouraging. For the indices of emerging markets, i.e. Budapest stock market, Warsaw stock market index, and Prague stock market, the heavy tail property of their returns can be reliably modeled by simple t-distribution, whose PDF, CDF and ICDF are known. For index ATX of the Vienna stock market, the skewed versions can be used. However, in this case, the search for a better candidate should not be excluded.

Acknowledgements

This paper is financially supported by grant SGS17/196/OKH4/3T/14.

References

[1] Ass, K., & Haff, I. H. (2006). The Generalized Hyperbolic Skew Student's t-Distribution. *Journal of Financial Econometrics*, 4(2), 275–309.

[2] Borak, S., Hardle, W., Weron, R. (2005). Stable Distributions. *SFB 649 Discussion Paper 2005-008*, Center for Applied Statistics and Economics, Humbolt-Universitat zu Berlin.

[3] Borak, S., Misiorek, A., Weron, R. (2010). Models for Heavy-tailed Asset Returns, *MPRA paper No. 25494*.

[4] Málek, J., Tran, V. Q. (2016). Application of alpha-Stable Distribution on Czech Financial Market. In: *Proc. of the 21th International Conference TPAPF 2016*, Praha: Oeconomica, 270–273.

[5] Málek, J., Tran, V. Q. (2017). Fat tail property of Czech financial assets. In: *Advanced Methods of Computational Finance*. Praha: Oeconomica.

[6] McDonald, J., & Newey, W. (1988). Partially Adaptive Estimation of Regression Models via the Generalized t Distribution, *Econometric Theory*, 4, 428–457.

[7] Nolan, J. P. (1997). Numerical Calculation of Stable Densities and Distribution Functions, *Communications in Statistics – Stochastic Models* 13, pp. 759–774.

[8] Prause, K. (1999). The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures, Ph.D. Thesis, Freiburg University.

[9] Snedecor, G. W., Cochran, W. G. (1989). *Statistical Methods* 8e. Iowa State University Press.

[10] Theodossiou, P. (1998). Financial Data and the Skewed Generalized t Distribution, *Management Science*, 44, 1650-1661.

Usage of Weights in MCDA Methods Based on DEA Results

Martina Kuncová[1], Jana Sekničková[2]

Abstract. In real economic analyzes we often come to a situation when it is necessary to compare selected units according to several criteria. It might be a situation of a comparison of countries, regions, municipalities, companies, products, etc. In such cases, one of the possibilities is the usage of Data Envelopment Analysis (DEA) to find efficient units, which in comparison with others achieve maximum outputs with minimum inputs. Similarly, multi-criteria decision analysis (MCDA) methods can be used to compare units according to more criteria. These methods, however, usually require the assignment of a weight vector that takes into account the importance of the criteria. Since one of the outputs of DEA models is the weight vector of input or output indicators, the aim of this article is to analyze the possibility of using these weights in selected methods of multi-criteria evaluation of alternatives. The analysis will be performed on real data to determine the similarity of DEA and MCDA results.

Keywords: data envelopment analysis, multi-criteria evaluation of alternatives, WSM, WSA, correlation.

JEL Classification: C44, C67
AMS Classification: 90B50, 90C29, 62H20

1 Introduction

Multi-criteria decision making (MCDM) belongs to operations research area as it uses various mathematical instruments, analysis and methods that should help the decision maker to find the best or the compromise solution or to better understand the reality. MCDM usually covers multi-criteria evaluation of alternatives methods sometimes called multi-criteria decision analysis (MCDA), but it can cover also multi-criteria linear programming [5] or other decision-making theories and concepts such as group decision-making or data envelopment analysis (DEA). When we thing about MCDA as the discrete decision-making problem, the main aim might be to find the good (or the best) alternatives according to several selected criteria [6]. Part of the MCDA methods uses cardinal information about the criteria in the form of criteria weights. Several approaches and methods are used to determine criteria weights. Usually weights are between 0 and 1 and they add up to 1. Weights can be set strictly by the decision-maker, or they can be calculated on the bases of the decision-maker preferences expressed by points or by pairwise comparison [11]. As it was mentioned above, DEA models are sometimes also taken as a tool useful for the decision-making process or for finding of the best alternatives out of the set of selected alternatives. In DEA analysis the alternatives are called decision-making units (DMU) and the main aim is to find out the efficient (best) units according to several criteria – here called inputs and outputs. One of the DEA results is also the weight vector that is calculated through an optimization model. Model searches for weights where the DMU would be efficient, i.e. effectively transforms its inputs into outputs. From this point of view the DEA and MCDA models and methods seems to be similar as they both use weights to find the best alternatives. MCDA need the weight vector to be described by the decision-maker but in DEA the weights are calculated. We decided to analyze whether it is possible to use DEA for the weights calculation and afterwards to use these weights in selected MCDA methods to obtain similar results, i.e. if the alternative (or DMU) is efficient in DEA model so it is also good (or best) in MCDA method.

In this paper we compare results taken from DEA models with results of basic MCDA methods when the average weights taken from DEA models are used. For the analysis data about several economic activity characteristics of 14 Czech regions were used.

2 Methods and data

According to [7] the DEA models and the MCDA methods using weights and calculation the utility for each alternative belong to the MCDA choice and ranking problems. In this part the basic description of DEA is made

[1] University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067, Prague 3, kuncovam@vse.cz
[2] University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067, Prague 3, jana.seknickova@vse.cz

followed by the short description of WSM, SAW and WSA methods as representatives of MCDA. Data for the analysis were taken from the Czech statistical office [4]. They describe the selected economic characteristics of the 14 Czech regions. They were used in the comparison of the Czech regions via other MCDA methods in [8].

2.1 DEA models

The basic idea of DEA models consists in estimation of an efficient frontier that defines production possibility set of the problem. DEA estimates so-called efficient frontier on the basis of the set of available decision making units (DMUs) and afterwards projects all DMUs onto this frontier. If a DMU lies on the frontier, it is referred to as an efficient unit, otherwise inefficient. This nonparametric method was initiated by Charnes, Cooper and Rhodes [1] and so the basic model is usually called CCR model according to the names of these authors. DEA models can be oriented to inputs or outputs depending on which indicators we understand as fixed. In the case of input oriented models we assume fixed level of all outputs (CCR-I) and the output oriented model assumes fixed level of all inputs (CCR-O) [1]. These models are used if we assume constant return to scale. In the case of variable return to scale, we work with BCC (Banker, Charnes, Cooper) models that can be similarly input or output oriented. The review and detailed information about DEA models are in [1] and [3]. The basic idea for the efficiency calculation is to maximize the rate of weighted sum of outputs divided by weighted sum of inputs [1].

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

$$TE_q = \frac{\sum_{j=1}^{n} u_j y_{jq}}{\sum_{i=1}^{m} v_i x_{iq}} = \frac{\mathbf{u}^{\mathrm{T}} \mathbf{y}_q}{\mathbf{v}^{\mathrm{T}} \mathbf{x}_q}, \tag{1}$$

where v_i, i = 1, 2, ..., m is a weight for i-th input, u_j, j = 1, 2, ..., n is a weight for j-th output and $\mathbf{u}, \mathbf{v}$ are vectors of these weights.

In this paper we work with all four basic DEA models (CCR-I, CCR-O, BCC-I and BCC-O) and so we use one universal DEA model (derived from original DEA models using Charnes-Cooper transformation) in the following form:

$$\begin{aligned} &\textit{maximize} \\ &\qquad z = i\left(\mathbf{u}^{\mathrm{T}} \mathbf{y}_q + b\omega\right) - (1 - i)\left(\mathbf{v}^{\mathrm{T}} \mathbf{x}_q + b\omega\right), \\ &\textit{subject to} \\ &\qquad i\left(\mathbf{v}^{\mathrm{T}} \mathbf{x}_q\right) + (1 - i)\left(\mathbf{u}^{\mathrm{T}} \mathbf{y}_q\right) = 1, \\ &\qquad \mathbf{u}^{\mathrm{T}} \mathbf{Y} + ib\omega \le \mathbf{v}^{\mathrm{T}} \mathbf{X} + (1 - i)b\omega, \\ &\qquad \mathbf{u} \ge \varepsilon, \\ &\qquad \mathbf{v} \ge \varepsilon, \\ &\qquad \omega - free. \end{aligned} \tag{2}$$

In the model (2) symbol ε denotes any small number, i and b denote parameters of universal DEA model ($i = 1$ for input oriented model and $i = 0$ for output oriented model, $b = 1$ for BCC model and $b = 0$ for CCR model). The efficient unit U_q lies on the efficient frontier in case that the optimal efficiency (calculated by the model) $|z| = 1$. The inefficient units have $|z|$ lower (in input oriented model) or higher (in output oriented model) than 1.

2.2 MCDA methods

Multi-criteria decision analysis (MCDA) methods have been developed to help the decision-maker in finding a solution among a lot of different alternatives on the basis of several criteria. MCDA methods can solve several types of problems, especially sorting problems, ranking problems and choice problems [7]. Main areas of the methods' usage can be found at [9]. As we talk about evaluation of alternatives, the term multi-criteria evaluation of alternatives is sometimes used instead of MCDA. As these methods use different principles and different information including subjective information they do not have the same results. To solve this kind of model it is necessary to know the preferences of the decision maker when all the alternatives (a_1, a_2, …, a_p) and criteria (f_1, f_2, …, f_k) are known. These preferences can be described by aspiration levels (or requirements), criteria order or by the weights of the criteria (usually stated as w_j for j = 1, .., k). For some methods it is enough to add this kind

of information but the more complex the method is the more inputs it needs. Because we want to compare the results with DEA models when the weights obtained by DEA are used by MCDA methods, we decided to start with the simple MCDA methods that do not need any additional information (except of weights). These methods are WSM, SAW and WSA. For all of them the formula (3), for the calculation of the final utility $u(a_i)$ for each alternative i, is common, the only difference is the calculation of r_{ij} (if w_j are the criteria weights).

$$u(a_i) = \sum_{j=1}^{k} w_j r_{ij}, \ \forall\, i = 1, \cdots, p \tag{3}$$

Weighted sum method (WSM) is one of the simple methods commonly used in single dimensional approach [11]. This method does not uses any criteria normalization function and for the calculation of the final utility of the alternatives it uses real data, so

$$r_{ij} = a_{ij}. \tag{4}$$

SAW (Simple Additive Weighting) method belongs also to the popular methods. It normalizes real data and several normalization formulas can be used [10]. In our analysis we calculate r_{ij} as in formula (5) for the maximization type of criterion or (6) for minimization type.

$$r_{ij} = \frac{a_{ij}}{\max_i a_{ij}} \tag{5}$$

$$r_{ij} = \frac{\min_i a_{ij}}{a_{ij}} \tag{6}$$

WSA (Weighted Sum Approach) is a case of SAW method where for the normalization different formulas are used [10] – instead of (5) it is formula (7) and instead of (6) it is (8).

$$r_{ij} = \frac{a_{ij} - \min_i a_{ij}}{\max_i a_{ij} - \min_i a_{ij}} \tag{7}$$

$$r_{ij} = \frac{\max_i a_{ij} - a_{ij}}{\max_i a_{ij} - \min_i a_{ij}} \tag{8}$$

2.3 Data used for the analysis

In our study we compared 14 Czech regions with respect to available data from 2016 using DEA and MCDA models (no.1 – Prague, the Capital City, no.2 – Central Bohemian Region, no.3 – South Bohemian Region, no.4 – Plzeň Region, no.5 – Karlovy Vary Region, no.6 – Ústí nad Labem Region, no.7 – Liberec Region, no.8 – Hradec Králové Region, no.9 – Pardubice Region, no.10 – Vysočina Region, no.11 – South Moravian Region, no.12 – Olomouc Region, no.13 – Moravian-Silesian Region and no.14 – Zlín Region). The models have 5 inputs (unemployment rate, economic activity, average wage, average age and free workplaces per capita) and 3 outputs (income per capita, consumption per capita and investments per capita). The DEA model assumes inputs are minimized and outputs are maximized for efficient units. In this case we had to transform input data to minimize them (especially economic activity and average wage). These data were described in [8] in details. For this paper we use this example as a case where more efficient units might be expected (because of higher number of inputs and outputs compared to the number of units).

3 Results

The presented analysis consists of three main steps. In the first step we use DEA models to obtain weight vectors. In the second step we use these vectors for MCDA methods. And in the third step we analyze correlation between DEA and MCDA results.

So firstly we solved all four presented DEA models by changing input parameters i and b in model (2). Note that variables in DEA model are weights of inputs (v_i) and outputs (u_j). Optimal solution for selected q-th unit found such weights that q-th unit is maximal efficient. This weights vector we take as an optimal weights for the unit $\boldsymbol{q}$ to have maximal efficiency score. It seems that using of this weight vector in MCDA models should provide very good rank for q-th unit in the case it is efficient. As we have 14 units we obtain 14 weights vectors for each DEA model – it means 56 weight vectors (for 4 DEA models).

In the second step we used these 56 weights vectors in each of three MCDA methods presented in this paper. Then we analyzed results with respect to DEA efficiency.

Finally we calculate correlation r between rank of DEA model and rank of MCDA model (as a measure of results dependency) for each weight vector and we use statistical t-test for null hypothesis $H_0: r = 0$ against alternative hypothesis $H_A: r \neq 0$. In this step we analyze also correlation between efficiency values gained from DEA and utility values gained from MCDA. When more efficient units were found by DEA model, all were

places into the same position calculated as the average rank of all efficient DMUs. To find the unique order the superefficiency model could be used but in this paper we use only basic DEA models.

3.1 The WSM results

For every weight vector from the first step we applied WSM method on real data (Czech regions). In both CCR models only two regions were efficient (no. 1 – Prague, the Capital City and no. 11 – South Moravian). Prague, the Capital City was placed on the first rank also by WSM by using of the optimal weight vector for the 1st unit. By using of the optimal weights for 11th unit South Moravian is placed on the first rank together with second efficient unit (Prague, the Capital City).

We obtain very similar results in the cases of BCC models. There are 7 efficient units (no. 1, 2, 6, 10, 11, 12, 14) and for each q-th optimal vector q-th unit is the winner or it is placed before all inefficient units.

We calculated all correlation coefficient between DEA rank and WSM rank (we use Pearson correlation as well as Spearman correlation) and we tested these values by t-test. The obtained correlations were generally high (absolute value more than 75 %) and only in three cases ($q = 1, 4, 9$) the correlation coefficient was non-significant at 5 % level.

For correlation between efficiency from DEA and utility from WSM the values were generally also high (absolute value more than 68 %) and only in the same three cases ($q = 1, 4, 9$) the correlation coefficient was non-significant at 5 % level.

3.2 The SAW results

The results for WSM were expected because DEA models set weight vectors for maximal efficiency that is measured as in (1) and the utility measured by (3) in WSM has similar formula. Contrariwise the results of SAW and WSA was slightly surprising because they were conformable to WSM results. Using of the optimal weight vector for the 1st unit places Prague, the Capital City on the first rank in cases of all DEA models.

SAW with the optimal weight vector for the 11th unit places unit no. 11 – South Moravian – at the third place behind the efficient Prague, the Capital City at the first place and inefficient no. 6 – Ústí nad Labem – at the second place. Note that Ústí nad Labem region had in CCR models efficiency almost 1 (0.975 in CCR-I, 1.026 in CCR-O) and it is efficient in both BCC models.

Again we calculated all correlation coefficients between DEA rank and SAW and we tested these values by t-test. The obtained correlations were high (absolute value more than 60 %) and only in three cases mentioned above ($q = 1, 4, 9$) the correlation coefficient was non-significant at 5 % level.

The correlation coefficients between efficiency from DEA and utility from SAW were generally high (absolute value more than 70 % for CCR models and 90 % for BCC-O) and slightly lower (more than 52 %) for BCC-I. At the 5 % level the three above mentioned cases had non-significant correlation in both CCR models and BCC-O model. In BCC-I model all correlation coefficients are non-significant at 5 % level, but only three above mentioned cases had non-significant correlation at 6 % level.

The unexpected results were obtained by using of the optimal weight vector for the 2nd unit in both BCC models. This unit was efficient, BCC models set optimal weights for maximal efficiency of this unit but SAW method using this weight vector placed the unit no. 2 – Central Bohemian – at the 10th place. First five places are occupied by efficient units, the next four places by inefficient units (that are not good) and after them efficient the Central Bohemian Region.

3.3 The WSA results

WSA was the third used method. The obtained results are very similar to the SAW method. For all DEA models Prague, the Capital City was placed as the first by using of the optimal vector for the 1st unit. WSA with the optimal weight vector for the 11th unit places unit no. 11 – South Moravian – at the third place behind the efficient Prague, the Capital City and inefficient no. 6 – Ústí nad Labem. The unexpected results in the case of the optimal 2nd weight vector have occurred also in the case of WSA method.

The obtained correlations between DEA rank and WSA rank were about 68 % for CCR models and 57-67 % for BCC models. Again three cases mentioned above ($q = 1, 4, 9$) had the correlation coefficients non-significant at 5 % level.

The correlation coefficients between efficiency from DEA and utility were generally high also for WSA (absolute value more than 68 and 71 % for CCR models and 69 % for BCC-O) and slightly lower (more than 51 %

for BCC-I). At the 5 % level the three above mentioned cases had non-significant correlation in both CCR models and BCC-O model. In BCC-I model all correlation coefficients are non-significant at 5 % level, but only three above mentioned cases had non-significant correlation at 6 % level.

3.4 The results with average weights

The most results of all three methods were very good but for the 1st, 4th and 9th weight vector for each DEA model correlations between DEA and MCDA results were non-significant. It can mean the results are independent. The goal of this analysis was to set a weight vector that provides the results by MCDA method very similar to DEA models. So we tried to calculate average weight vector as the mean of all (in our case fourteen) optimal weight vectors. Then we tried to use its components as weights w_j in formula (3). The results were amazing. The ranks obtained by WSM was almost the same as in DEA models, the ranks for SAW and WSA were only slightly different. We calculated correlation coefficients between DEA rank and MCDA methods rank and we tested these values by t-test. The calculated correlations and p-values for t-test are displayed in the left part of Table 1. From this table we can see that for all methods and all DEA models the results gained by using average weight vector in MCDA methods are dependent on the results gained by DEA at 3.5 % level and each correlation coefficient is at least 57.5 %.

		Rank correlation			**Utility correlation**		
		WSM	**SAW**	**WSA**	**WSM**	**SAW**	**WSA**
CCR-I	correlation	0.9813	0.6821	0.6821	0.9382	0.7037	0.6811
	p-value	0.0000	0.0072	0.0072	0.0000	0.0050	0.0073
BCC-I	correlation	0.7252	0.5750	0.5750	0.7224	0.5255	0.5145
	p-value	0.0033	0.0315	0.0315	0.0035	0.0536	0.0598
CCR-O	correlation	0.9813	0.6821	0.6821	0.9475	0.7223	0.7135
	p-value	0.0000	0.0072	0.0072	0.0000	0.0035	0.0042
BCC-O	correlation	0.9083	0.6689	0.6689	0.9085	0.6876	0.6936
	p-value	0.0000	0.0089	0.0089	0.0000	0.0066	0.0059

Table 1 Correlation between DEA and MCDA for average weights

The similar results we obtained in the case of correlation between efficiency from DEA and utility from MCDA methods. The values of correlation coefficients and p-values for t-tests are displayed in the right part of Table 1. From this table we can conclude that all correlation coefficients are higher than 51 % and they are significant at 6 % level (except BCC-I models even at 0.9 % level).

3.5 The results with normalized average weights and average of normalized weights

For all previous analyses we used weights in the form gained directly from model (2). As MCDA models usually works in formula (3) with normalized weights (i.e. $\sum_{j=1}^{k} w_j = 1$) we tried to normalized average weights from previous analysis. The correlation between DEA and MCDA ranks with corresponding p-values are obviously the same as displayed in Table 1.

In the last case we used also average of normalized weights gained from DEA models. That means we took 14 weight vectors from DEA as written in chapter 3 and we normalized each of them. Then we calculated average of these 14 normalized weight vectors. This vector we used for MCDA analysis.

From the Table 1 we can see, for example, that the correlation for average weight vector (and also for normalized average weight vector) obtained from CCR-I model and used in WSA method is about 68 % for the rank as well as for the utility and it is significant at 0.75 % level. If we use average of normalized weights instead of normalized average weights the rank correlation is 71.11 % with p-value 0.44 % and the utility correlation is 77.43 % and it is significant at 0.12 % level. Such results are much better and this approach should be a matter of the next research.

4 Conclusions

The aim of our analysis was to find out whether the weights gained from DEA models can be used in selected MCDA methods to obtain identical evaluation of variants in terms of effective/good/bad order position. The analysis and the results comparison showed that only in the case of the average weights taken from DEA models (or the average normalized weights) the results of MCDA methods are highly correlated with DEA results. These results are practically independent on the MCDA methods used or on the DEA model applied. The position of the units (here regions) that were efficient in DEA models is also (with exceptions) good in MCDA models. When we use the weight vectors obtained in the DEA model for each of the unit (region) separately, then we can no longer count on similar ratings, and some efficient units can place worse in MCDA method evaluation than units inefficient. In the next research we would like to verify results of this paper on other data file and by using additional MCDA methods to see, whether using of DEA weights in MCDA methods is meaningful or not.

It is clear that the results of CCR and BCC models – especially in the case of small number of DMUs with higher number of inputs and outputs – provide more efficient units. In such a situation, it would be more appropriate to use superefficiency models which will determine the exact order of units. This will be also a matter of further research.

Acknowledgements

The research was supported by the grant No. 16-01821S of the Czech Science Foundation and also by the Internal Grant Agency of the University of Economics in Prague IGA F4/57/2017.

References

[1] Charnes A., Cooper W.W. & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, pp. 429-444.
[2] Cooper, W. W., Lawrence, M.S. & Zhu, J. (2004*). Handbook on Data Envelopment Analysis*. Norwell: Kluwer Academic Publishers.
[3] Cooper, W.W., Seiford, L.M. & Tone, K. (2006). *Introduction to Data Envelopment Analysis and Its Uses*. New York: Springer Science + Business Media Inc.
[4] Czech Statistical Office, (2018) [online], available at https://www.czso.cz/csu/czso/ [Accessed 17 Jan 2018]
[5] Fiala, P. (2006). *Modely a metody rozhodování*. Praha: Oeconomica.
[6] Figueira, J., Greco, S. and Ehrgott M. (2005). Multiple Criteria Decision Analysis - State of the Art Surveys. New York : Springer Science + Business Media Inc.
[7] Ishizaka, A., Nemery, P. (2013). *Multi-Criteria Decision Analysis*. UK: John Wiley & Sons, Ltd.
[8] Kuncová, M., Sekničková, J. (2018). Multi-criteria evaluation of the Czech regions from the selected economic activity aspects point of view. *Quantitative Methods in Economics (Multiple Criteria Decision Making XXI)*, Trenčianske Teplice. Forthcoming.
[9] Mardani, A., Jusoh, A., Nor, K.MD., Khalifah, Z., Zakwan, N. & Valipour, A. (2015). Multiple criteria decision-making techniques and their applications – a review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istraživanja,* 28(1), pp. 516-571.
[10] Podviezko, A., Podvezko, V. (2015). Influence of data transformation on multicriteria evaluation results. *Procedia Engineering* 122, pp. 151-157.
[11] Triantaphyllou, E. (2000). Multi-Criteria Decision Making Methods: A Comparative Study. Dordrecht: Kluwer Academic Publishers

Prices Indices and Inflation Measurement: Scanner Data Challenge

Hana Lipovská [1] Jakub Fisher [2]

Abstract. The Great Recession of 2008, connected to extremely low inflation rates and Zero Lower Bound, resulted in the unconventional monetary policy of central banks. Moreover, shortly after the end of the bust, several national statistical bureaus started to use completely new method of collecting data on consumer prices. The aim of this contribution is to explain how the calculation of the Consumer Price Index changed with scanner data introduction. We summarise the key features of the scanner data and test hypothesis on change of the inflation ratio both in terms of mean and volatility after the implementation of Scanner Data Approach in Belgium and Denmark in comparison to their pair countries (France and Finland). We found that there was no structural change in Danish inflation time series, but there was upward shift in Belgium time series even in comparison to France. Our results can contribute to the current discussion on the introducing the scanner data by the Czech Statistical Office.

Keywords: Consumer Price Indices, CUSUM test, Missing Inflation, Scanner data.

JEL Classification: C22, C82, E31
AMS Classification: 62-07

1 Introduction

There are two macroeconomic indicators which are powerful enough to affect elections: inflation rate and unemployment rate. While unemployment rate is comprehensible for everyone on the labour market, inflation is understood through the consumer's costs of living. Being *"always and everywhere a monetary phenomenon"* [5], it has been since 1970's the main objective of the leading central banks. In 1974 inflation was declared "Public Enemy No. 1" by President Gerald Ford [9]. Forty years later, the main objective remains the same, but the circumstances changed substantially. Immediately after the Great recession the *"Mystery of the missing inflation"* had occurred [10]. The average inflation rate in Eurozone from the beginning of the common euro currency in 1999 to the beginning of the Great Recession was 2.2%, however it failed to merely 1.2% in the following period 2009–2017.

When trying to explain the *Mystery of the missing inflation*, new unconventional monetary policies are being discussed. This paper on the contrary tests the impact of new statistical methods. The inflation rate is estimated using the Consumer price index (CPI). The first CPI was constructed for the Czechoslovak yet in 1920 by Mildschuch. Originally the basket had contained 11 industrial products; however it was quickly spread to 69 products by 1922. The CPI has been calculated since 1921 for the consumer basket of 25 goods bought by the typical five-member working class family and four-member clerical family [4]. Nowadays, the consumer basket convers about 640 items which prices are collected monthly

The main weaknesses of such approach are well known. As Boskin Comission reported in 1996, the inflation rate tends to be upward biased by 1.1 % p. a. [6] because of the substitution of the products, outlet substitution, new products introduction or quality change. Nevertheless, in the last decades the technical growth as well as the development of Big Data processing has allowed to use new method of Consumer Price Index calculation: the Scanner Data Approach [3]. Scanner Data Approach is used e. g. in Dutch, Norway, Switzerland, Sweden, Belgium, Denmark or Australia, the Czech Statistical Office is going to test it in 2018 [1].

According to the recent findings, the scanner data may contribute to the lower inflation rates, as they cope with the main weaknesses of the traditional Shelf Prices Approach (e.g. Watanabe and Watanabe [13]). Furthermore, the inflation rate volatility tends to be three times higher when estimated by the CPI based on the Scanner Data Approach in comparison to the Shelf Prices Approach [7]. In this contribution we test those findings on the European data published by Eurostat.

The rest of the paper is organized as follows. Firstly, we will briefly summarise the key features of the scanner data. Then we describe dataset and econometric methods used. In the third part two hypotheses are tested.

[1] Masaryk University, Faculty of Economics and Administration, Lipová 41, Brno, lipovska@mail.muni.cz
[2] University of Economics, Department of Economic Statistics, W. Churchila 4, Prague, fisherj@vse.cz

With the Scanner Data Approach we stand above the cradle of new exiting method, which might improve our knowledge of economy as well as new statistical and econometrical technics. On the other hand, it means also revolution in the standard procedures which can seriously influence monetary policy. That is why it must be studied very carefully from the very beginning.

2 Scanner data

Scanner data are detail data on the prices and trade volume of consumer good based on the bar codes scanned in the supermarkets [3]. Since 1974, when the bar code had been implemented, to the beginning of the millennium, the scanner data were used mainly for the marketing purposes. The scanner data usage for the aim of inflation measurement was discussed later [11].

The main advantage of the scanner data is the huge market coverage. If the statistical authority makes an agreement with the chain stores, it receives detailed data on the prices, volumes and qualities not for 600 or 700 but for 10 000 –25 000 items on the whole market [3]. *Statistics Norway* covers 96% of the chain stores, 100% of the alcohol stores, 85% of the pharmacies as well as petrol stations [8].

Moreover, the traditional Shelf Prices Approach (SPA) is based on the data collection only in the certain working days and it is demanding on the human capital, time as well as financial sources. That is why the price data are available only for the larger towns. Shelf prices also cannot mirror the real price paid by the customer: they do not include the special offers for the club members, sales, promotions etc. Scanner Data Approach (SDA) do not work with one price p_i for the *i*-th item, but it calculate average price $\bar{p}_i$ based on the total expenditures on the *i*-th item TE_i and quantity sold Q_i

$$\bar{p}_i = \frac{TE_i}{Q_i} \tag{1}$$

Especially states with higher inflation rate (e.g. in Latin America) can benefit from the prompt availability of data with minimal lag; inflation rate can be estimated also with higher frequention (weekly or even daily, [13]).

Scanner data can be applied in static or dynamic way. Static method just replicate the standard Consumer Price Index based on the fix weights and calculated as Laspeyres price index. On the contrary, the dynamic method allows cover even the new products in real time. For the aim of dynamic computation the bilateral index based on Geary-Khamis method is used [2].

The main weakness of the scanner data lies in the legislative conditions (as the special agreements between statistical authorities and private companies has to be made), as well as in its technical aspects. There is certain risk of not delivering data in time (because of the negligence or reluctance of companies or technical obstacles). Moreover, scanner data are Big-Data in kind, which is demanding on the IT and new statistical methods.

Potential risk lies in the change of the data collection method from SPA to SDA. If the SDA method affects volatility or level of inflation, it can make the efficient monetary policy harder because of its long lags (12–18 months for the Czech National Bank). That´s why the change in inflation rate level and volatility after SDA implementation is subject of our research.

3 Data and Methodology

For the aim of our research, Eurostat data on the inflation rate were used. Monthly inflation rate is calculated as the annual rate of change of the *Harmonised index of consumer prices* covering all items. We used data for period from January 1999, when the Eurozone was founded, to the March 2018 for which the last data are available.

Two pairs of countries were connected according to following criteria:

- All countries must be member states of the EU.
- Times series of the inflation rate in the whole period should be strongly correlated for countries in each pair.
- One of the countries in each pair is using Scanner Data Approach, while the second one uses wholly the Shelf Prices Approach.

The correlation matrix was estimated for 15 old EU member states. The new EU member states – especially the Central and Eastern European countries – were excluded from this analysis, as their inflation rate was on the beginning of the period still influenced by the economic transformation. Based on these criteria, the following pairs of countries were chosen:

a) Belgium and France: correlation of their inflation rate is very strong and statistically significant ($\rho = 0.80, p < 0.05$). Belgium fully introduced the Scanner Data Approach in January 2015.

b) Denmark and Finland: correlation between inflation rate is very strong and statistically significant ($\rho = 0.76, p < 0.05$). Denmark fully introduced the Scanner Data Approach in January 2016.

The monthly inflation rates are subject of the fluctuations, the time series were therefore smoothed using the centred moving average of then length = 13:

$$\bar{y} = \frac{(y_{t-6} + 2y_{t-5} + 2y_{t-4} + \cdots + 2y_t + 2y_{t+1} + \cdots + 2y_{t+5} + y_{t+6})}{24} \quad (2)$$

The smoothed time series (Figure 1) are used to find the structural breaks caused by shocks (e. g. recessions etc.)

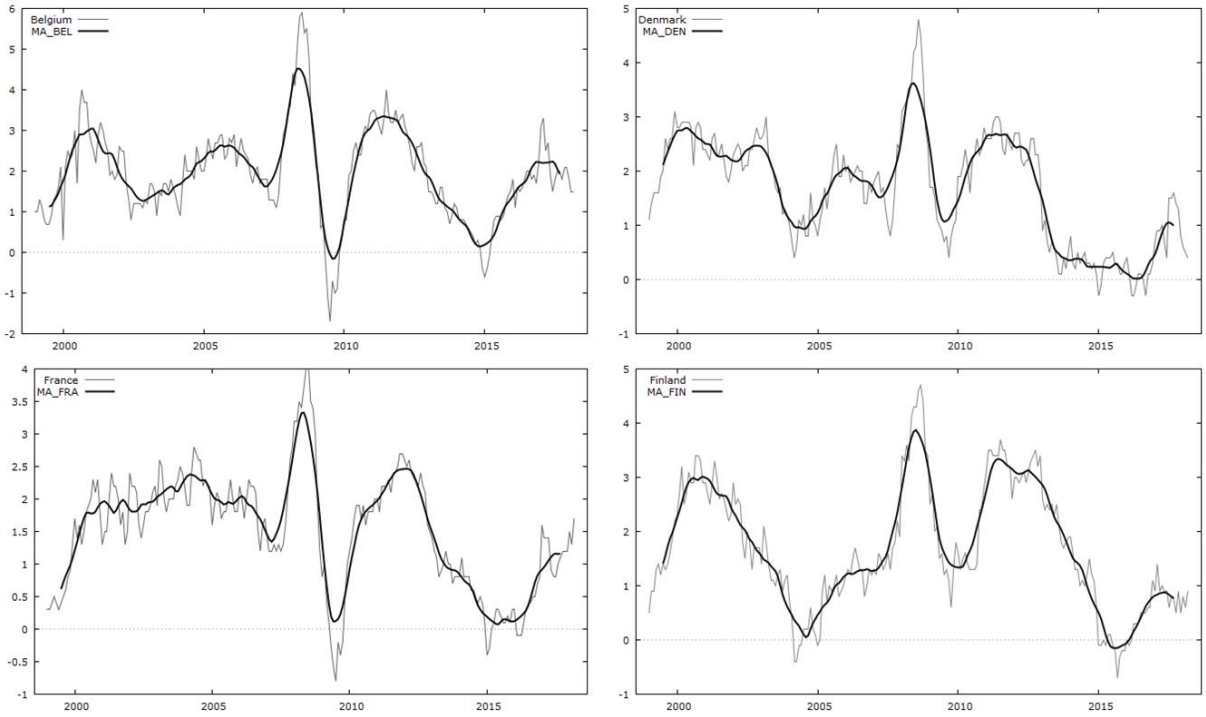


Figure 1 Inflation rate time series smoothed by the centered moving average

Based on these time series we test two theoretical hypotheses:

H1: Inflation rate decreases if country starts to use the Scanner Data Approach.
H2: Volatility of inflation rate increases if country starts to use the Scanner Data Approach.

To test hypotheses H1 and H2, tests of the structural break were used. CUSUM test for structural change is based on the cumulative sums of deviations of the sample values from the target values [12]. Similarly the CUSUMSQ test allows testing the homoscedasticity.

CUSUM test helps us to detect, where the structural break could be. However, it does not help us in deciding if the potential structural break might be result of the swap to Scanner Data Approach, or if the change in the average inflation rate or inflation variability is caused by some latent factor. The simple OLS model (3) was therefore estimated to test the structural change by CUSUM test and Chow stability test with respect to the pair country:

$$\pi_{SD} = \beta_0 + \beta_1 \pi_{SP} + \varepsilon, \quad (3)$$

where π_{SD} is the inflation rate for country using Scanner Data Approach (Belgium and Denmark), while π_{SP} is the inflation rate for country using Shelf Prices Approach (France and Finland).

Furthermore, two-sample t-test for comparing two means was used to test the null hypothesis:

$$H_0: \mu_1 = \mu_2 \quad (4)$$

The F-test was used to test the equality of two variances:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad (5)$$

For the aim of our research, the econometric software Gretl 1.19.14 and statistical software Dell Statistica version 13 were used.

4 Results

In this chapter theoretical hypothesis H1 and H2 are tested. Firstly, the change of mean and variability of the inflation rate time series for Belgium, France, Denmark and Finland are tested. Then the CUSUM test is used for the simple naive regression of the inflation rate to the constant to identify the periods of the homogeneous inflation rate. Finally, the regression of the inflation rate based on Scanner Data Approach to the time series of paired country is calculated and the potential structural breaks are tested with the Chow test.

Table 1 summarise the mean of inflation rate as well as the standard deviation before the Scanner Data Approach was implemented in Belgium (2015) and France (2016) and after the implementation. The inflation rate decreased in all four countries, this decrease is statistically significant on the level of significance 5 %.

	1999–2018		**1999–2014**		**2015–2018**		$\boldsymbol{H_0: \mu_1 = \mu_2}$
	Mean	st. Dev.	Mean	st. Dev.	Mean	st. Dev.	p-value
Belgium	1.95	1.18	2.03	1.22	1.55	0.82	0.033
France	1.51	0.91	1.69	0.84	0.59	0.58	1.28 e-011
	1999–2018		**1999–2015**		**2016–2018**		$\boldsymbol{H_0: \mu_1 = \mu_2}$
	Mean	st. Dev.	Mean	st. Dev.	Mean	st. Dev.	p-value
Denmark	1.65	1.02	1.80	0.98	0.53	0.58	4.947e-006
Finland	1.69	1.14	1.83	1.14	0.63	0.35	3.773e-008

Table 2 Average inflation rate and its variance for

Structural breaks in the individual time series were identified by the CUSUM test. Furthermore, we teste the structural breaks also in the shorter times series smoothed by the centred moving average (Table 2). The significant structural breaks occurred during the Great Recession. For the aim of our research, we restricted the time series to the sub-period March 2009 – March 2018 for Belgium and France and August 2008 – March 2018 for Denmark and Finland.

		CUSUM breaks				
original time series	Belgium	7/2001	6./2008	6/2009	11/2011	4/2013
	France	6/2001	3/2014			
	Finland	4/2001	8/2002			
	Denmark	1/2015				
centred moving average	Belgium	10/2008	3/2009			
	France	9/2009	10/2016			
	Finland	7/2008	7/2017			
	Denmark	7/2008	1/2010	11/2013		

Table 2 Structural breaks in time based on CUSUM test

Linear regression (3) is estimated using additional regressor $D_i, i \in \{1; 2\}$. Regressor D_1, specific for Belgium, is the dummy variable of value 1 for period from September 2009 to March 2018 and value 0 otherwise. Regressor D_2 , used for Denmark, is the dummy variable of value 1 for period from September 2015 to March 2018 and 0 otherwise. Results of both OLS models are summed up in Table 3.

Model 1: Belgium				
	Coeff.	**St. Dev**	**p-value**	
const	0.29	0.22	0.19	
France	1.16	0.13	3.63e-015	***
D1	0.49	0.21	0.0234	**
R2				0.74
Model 2: Denmark				
const	-0.07	0.24	0.78	
Finland	0.76	0.09	1.22e-014	***
D2	0.18	0.25	0.48	
R2				0.67

Table 3 OLS models estimates

Model 1 confirms the close correlation between French and Belgium inflation rate, however it also confirms, that there is shift in data after implementation of the Scanner Data Approach. Chow test confirms the structural break in September 2009 as suggested by Chow test (with *p*-value = 0.0001). Nevertheless, this shift in data is not consistent with our hypothesis H1 as the inflation rate in Belgium is 0.49 percentage points higher after the implementation of SDA. Increase in Belgium inflation rate after SDA implementation is confirmed also by *t*-test, while there is no significant change in France (Figure 2). On the other hand while Belgium inflation rate decreased significantly between period 2009:3–2016:8 and 2016:9–2018:3 (*t*-test rejects hypothesis on the equal means, *p*-value = 0.00052), the French inflation rate did not change at all (*p*-value = 0.89). For Denmark (Figure 3), there is no structural change between August and September 2015 detected relatively to the Finland (Chow test does not reject the hypothesis on the *p*-value = 0.39).

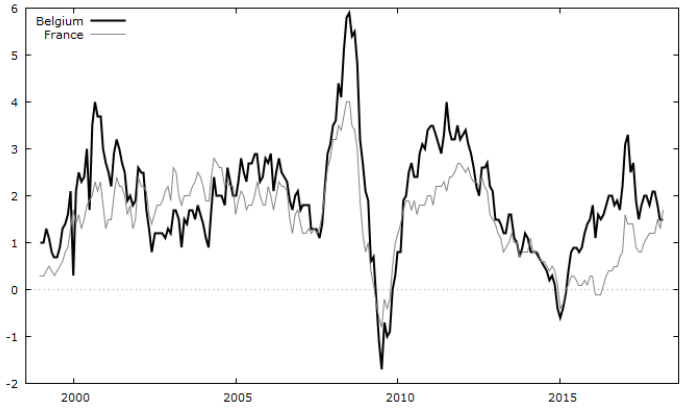


Figure 2 Inflation rate development in Belgium and France (1999–2018)

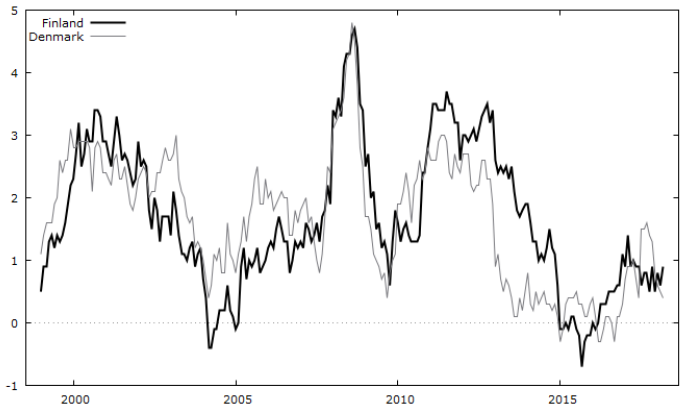


Figure 3 Inflation rate development in Finland and Denmark (1999–2018)

CUSUMSQ test confirms break in variability for Belgium relatively to France since March 2016, but it does not confirm any break in variability for Denmark relatively to Finland. Nevertheless, the variability of inflation rate in all four countries decreased significantly between both periods (p-value of F-test for equality of two variances is lower than 0.01).

5 Conclusion

Importance of the Scanner Data increases rapidly with the technical development of data collection and procession. Implementation of Scanner Data Approach in inflation rate estimates might help to cope with several traditional weaknesses of nearly one hundred years used CPI Shelf Prices Approach. Nevertheless, Scanner Data Approach is according to the recent findings from the Japanese economy connected to significant decrease of inflation rate together with significant increase in its volatility. The aim of this paper was to test hypothesis of inflation rate decrease and volatility increase on the two pairs of European data. Paired time series made the comparison available, as the potential downward structural break e. g. in Denmark data could have been caused both by implementation of SDA as well as by other, latent factor, which affected the "Mystery of Missing Inflation". That is why we compare the SDA data with SPA data of countries, which inflation rate is closely correlated to their counterparts.

There is significant structural break in the Belgium inflation rate as well as its volatility (in comparison to French data) after implementation of the Scanner Data Approach in 2015. However, contrary to the hypotheses H1 and H2, the inflation rate increased and the volatility decreased. For Denmark in comparison to Finland there was no significant change in inflation rate, nor its volatility.

According to our results, the changes in the inflation rate estimates cannot be fully attributed to the change in the data collection from Shelf Prices Approach to Scanner Data Approach. For the aim of the SDA implementation by the Czech Statistical Office it is important, that change in approach does might not seriously affect inflation rate stability. Nevertheless, according to Eurostat [3] it is important to prepare the implementation in close cooperation with the central bank as well and the academic community and media to avoid unfavourable shocks.

Acknowledgements

This paper was supported by Masaryk University, project No. MUNI/A/0806/2017.

References

[1] Báčová, P. (2018). The year 2018 will bring many challenges. [Rok 2018 přinese mnoho výzev] *Statistika a my*. ČSÚ.
[2] Chessa, A. G. (2016). *A new methodology for processing scanner data in the Dutch CPI.* Eurostat Review on National Accounts and Macroeconomic Indicators. Eurostat.
[3] Eurostat. (2017). *Practical Guide for Processing Supermarket Scanner Data*.
[4] Fousek, F. (1927). Economic Policy, Vol. 1 [Hospodářská politika, Svazek 1]. Vyd. František Fousek.
[5] Friedman, M. (1956). The Quantity Theory of Money-A Restatement. *Studies in the Quantity Theory of Money*, Chicago: University of Chicago Press, pp. 1–21.
[6] Gordon, R. J. (2006). *The Boskin Comission report: A retrospective one decade later.* NBER WP 12311.
[7] Imai, S., Watanabe, T. (2015). *Replicating Japan´s CPI Using Scanner Data*. WP No. 072.
[8] Johannessen, R. 2016. *Scanner data in CPI/HICP*. Statistics Norway.
[9] Öner, C. (2017). Inflation: Price on the Rise. *Finance and Development*, IMF.
[10] Roubini, N. (2013). *Mystery of the Missing Inflation*. Project Syndicate.
[11] Silver, M., Heravi, S. (2001). *Scanner data and the measurement of inflation*. The Economic Journal, 111(June). Royal Economic Society.
[12] Wachs, S. (2010). *What is a CUSUM Chart and When Should I Use One?* Integral Concepts.
[13] Watanabe, K., Watanabe, T. (2014). *Estimating Daily Inflation Using Scanner Data: A Progress Report*.

Multi-asset options with different payoff functions

Ladislav Lukáš[1]

Abstract. The paper deals with formulation of multi-asset option pricing problems with different payoff functions. Multi-variability is important concept in financial engineering as many non-standard structured products in the market are exposed to multiple source of randomness. Spot prices of underlying asset are assumed 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. Most payoff functions are assumed to be non-negative convex function over a convex domain depending on underlying asset prices. Some typical examples of them are presented. Within a framework of multi-asset options a pricing of rainbow trend options was presented recently. In general, they are desirable to investors due to their diversification effects over different assets and time. We present the numerical implementation of pricing these options in Mathematica.

Keywords: multi-asset options, payoff functions, option pricing, rainbow trend option, numerical implementation in Mathematica.

JEL classification: G13
AMS classification: 91G80

1 Introduction

Pricing financial derivatives is a vital and practical issue because their use for hedging is highly demanded in the financial markets in general. Both industry and academia are keen to discover the fair prices of financial derivatives by combining financial models with mathematical theories. There is well kown that options are financial contracts giving their owner the one-time-use right, but not obligation, to buy or to sell a portfolio of asset shares for a pre-agreed price, called exercise/strike price K, at a future time. The last valid time to exercise such a right is called expiry/maturity time, denoted T. The pricing problems of financial derivatives generally starts from modelling the dynamics of an underlying asset in terms of a discrete or continuous stochastic process, in general.

First, we make a short overview of multi-asset European option pricing problem formulation assuming spot prices of underlying assets follow geometric Brownian motions 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. For pricing purposes within this approach, one assumes that corresponding financial market is complete, i.e. it consists from m risky underlying assets and one standard risk-free asset, transactions are efficient and free of transactional costs and taxes. Such approach is discussed in [1], [2], [3], and [6], with different level of details and abstraction.

We consider m risky assets A_i with their spot prices $S_i(t), i = 1, \ldots, m$, 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

$$\mathrm{d}S_i(t) = S_i(t)(\mu_i \mathrm{d}t + \sigma_i \mathrm{d}W_i(t)), \quad i = 1, \ldots, m,\ t \geq 0. \tag{1}$$

where μ_i specify expected rate of returns (constants), being usually expressed by $r - q_i$, with r risk-free rate, and q_i dividend rates, σ_i are volatilities (positive constants), and $\{W_i(t)\}, i = 1, \ldots, m$ are correlated Brownian motions, or Wiener processes, with matrix of correlation factors ρ_{ij}, $-1 \leq \rho_{ij} \leq 1,\ i, j = 1, \ldots, m$, where ρ_{ij} is correlation factor (constant) between $W_i(t)$ and $W_j(t)$, thus giving the covariance $\mathrm{Cov}(S_i(t), S_j(t)) = \rho_{ij}\sigma_1\sigma_2$, between $S_i(t)$ and $S_j(t)$, and finally

$$\mathrm{d}W_i(t) \sim N(0, \mathrm{d}t), \quad \text{i.e. } \mathrm{E}(\mathrm{d}W_i(t)) = 0, \quad \mathrm{Var}(\mathrm{d}W_i(t)) = \mathrm{d}t, \quad \mathrm{Cov}(\mathrm{d}W_i(t), \mathrm{d}W_j(t)) = \rho_{ij}\mathrm{d}t.$$

Let $V(S_1, \ldots, S_m, t)$ be a multi-asset option price compounded from S_i, at time t, where we write simply S_i instead of $S_i(t)$, 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 is constructed by Δ-hedging principle using Δ_i shares of assets A_i having prices S_i at time t, thus taking the well-known form: $V(S_1, \ldots, S_m, t) - \sum_{i=1}^{m} \Delta_i S_i$.

[1]University of West Bohemia, Faculty of Economics, Dept. of Economics and Quantitative Methods, Univerzitní 22, 306 14 Pilsen, lukasl@kem.zcu.cz

Application of Itô's formula upon $\mathrm{d}V$, Δ-hedging principle yielding $\Delta_i = \frac{\partial V}{\partial S_i}, i = 1, \ldots, m$, and the arbitrage-free argument, leads finally to the Black-Scholes partial differential equation (B-S PDE) for multi-asset option price V

$$\frac{\partial V}{\partial t} + \mathcal{A}_m V - rV = 0, \quad \forall (S_1, \ldots, S_n, t) \in \mathbb{R}_+ \times \cdots \times \mathbb{R}_+ \times [0, T), \tag{2}$$

where the differential operator $\mathcal{A}_m$is given by its application on any function $v(S_1, \ldots, S_m, t)$ of $m+1$ variables from the class $C^{2,1}(\mathbb{R}_+^m \times [0, T))$ in classical sense

$$\mathcal{A}_m v = \sum_{i=1}^{m} (r - q_i) S_i \frac{\partial v}{\partial S_i} + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \rho_{i,j} \sigma_i \sigma_j S_i S_j \frac{\partial^2 v}{\partial S_i \partial S_j}. \tag{3}$$

For more technical details, we refer to [1], in particular. However, the problem formulation is discussed in [5], too. In order to conclude the multi-asset option pricing problem, we need to add a payoff function, denoted $g(S_1, \ldots, S_m; K)$, to Eq.(2). The payoff refers to the monetary value gained from exercising an option which depends on the exercise price K and underlying asset prices $S_1(T), \ldots, S_m(T)$ attained at maturity T. The function $g(S_1, \ldots, S_m; K)$ can take various forms and can be quite complicated. However, most of them are assumed to be non-negative convex functions over a convex domain.

2 Different payoff functions for multi-asset option pricing problems

There is well-known that multi-asset option composed from m underlying assets $S_i, i = 1, \ldots, m$ is called traditionally as

- *basket option*: $S^p(t) = \sum_{i=1}^{m} w_i S_i(t)$, with weights $w_i \geq 0, \sum_{i=1}^{m} w_i = 1$, as usual,
- *maximum option*: $S^p(t) = \max\{S_i(t)\}$, $i = 1, \ldots, m$,
- *minimum option*: $S^p(t) = \min\{S_i(t)\}$, $i = 1, \ldots, m$,

where $S^p(t)$ denotes a portfolio composed from $\{S_i\}$, at time t, in general.

When option pricing problems are considered from the aspect of B-S PDE, the proper terminal condition (TC) and boundary conditions (BCs) are requested for solving them. The TC specifies that the option value at T equals its payoff value, i.e.

$$V(S_1, \ldots, S_n, T) = h_P(S_1, \ldots, S_m; K), \ \forall (S_i, \ldots, S_m) \in \mathbb{R}_+^m. \tag{4}$$

Following [2], Chapter 1, we can define the following payoff functions being in correspondence with the type of multi-asset option

- for *basket call option*: $g(S_1, \ldots, S_m; K) = \max(K - \sum_{i=1}^{m} w_i S_i(T), 0)$, $w_i \geq 0, \sum_{i=1}^{m} w_i = 1$, as usual,
- for *maximum call option*: $g(S_1, \ldots, S_m; K) = \max(K - \max_{i=1,\ldots,m}\{S_i(T)\}, 0)$,
- for *minimum call option*: $g(S_1, \ldots, S_m; K) = \max(K - \min_{i=1,\ldots,m}\{S_i(T)\}, 0)$.

Note: For put options, the components within the payoff functions interchange their positions, only.

In case of multi-asset option pricing problem, the BCs are very important since we need to discuss behaviour of $V(S_1, \ldots, S_m, t)$ at the boundary of admissible domain $\Omega = \mathbb{R}_+^n \times [0, T)$. There are 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

$$\lim_{S_i \to +\infty} V(S_1, \ldots, S_n, t) = 0, \quad i = 1, \ldots, n, \ \forall t \in [0, T). \tag{5}$$

In case of numerical computation, we truncate the values of S_i by sufficiently large $S_{\max}$, each, thus replacing the domain Ω by truncated one $[0, S_{\max}]^m \times [0, T)$, i.e. with m-dimensional hypercube for state variables. However, we assume the similar results to hold for $S_{\max}$ as doing for $S_i \to +\infty$.

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 $m - 1$ asset prices S_i simultaneously. In order to illustrate such influence, we discuss a simple case of two-asset options, i.e. $S^p = \{S_1, S_2\}$. We also adopt notation $(g)^+ \equiv \max(g, 0)$ to extract non-negative values of g, as usual. In case of *basket call option*, we have

$$((w_1 S_1 + w_2 S_2) - K)^+ = \begin{cases} (w_2 S_2 - K)^+ = w_2 (S_2 - K/w_2)^+ & \text{on } S_1 = 0, \\ (S_{\max} - K) + w_2 S_2 >> 0, S_{\max} >> K & \text{on } S_1 = S_{\max}. \end{cases}$$

In case of *maximum call option*, we have

$$(\max(S_1, S_2) - K)^+ = \begin{cases} (S_2 - K)^+ & \text{on } S_1 = 0, \\ (S_{\max} - K) >> 0, S_{\max} >> K & \text{on } S_1 = S_{\max}. \end{cases}$$

In case of *minimum call option*, we have

$$(\min(S_1, S_2) - K)^+ = \begin{cases} (0 - K)^+ = 0 & \text{on } S_1 = 0, \\ (S_2 - K)^+, S_{\max} >> K & \text{on } S_1 = S_{\max}. \end{cases}$$

3 Rainbow trend/average option – RTO

Another type of multi-asset options was introduced recently, see [7]. From a theoretical point of view, RTOs are desirable to investors due to their diversification effects over different assets and time. However, maybe due to the complexity of combining the pricing techniques for RTOs, they are not explored in literature yet. The motive for studying RTOs is their ability to simultaneously resolve the issues of how to choose a better-performing asset and how to time the market, which are two central concerns for all investors. It is well known that rainbow options, defined as options with payoffs involving several asset prices, can solve the asset selection problem, especially if the maximum or minimum payoff function is considered. In addition, trend options can make the timing decision less importamt by linking the payoffs with the average trend of the asset price instead of the final value of the asset price at maturity.

The martingale pricing method is used in [7] to unify the pricing models for rainbow and trend options considering a highly general framework for the payoffs of RTOs thus providing rather general pricing formulae. We consider RTO on m risky underlying assets. In martingale pricing technique, the risk neutral probability measure plays crucial role. In contrast to the PDE approach, the martingale pricing technique is based upon conditional expectations. Therefore, we rewrite the Eq.(1) in order to emphasize that probability measure in following way

$$\mathrm{d}S_i(t) = S_i(t)\big((r - q_i)\mathrm{d}t + \sigma_i \mathrm{d}W_i^Q(t)\big), \quad i = 1, \ldots, m,\ t \geq 0. \tag{6}$$

where $W_i^Q(t)$ is the standard Brownian motion for the i-th underlying asset under the risk neutral probability measure Q, r is the risk-free interest rate, q_i and σ_i are the dividend yield and the price volatility of the i-th asset, respectively. The $W_i^Q(t)$ and $W_j^Q(t)$ are correlated with the coefficient ρ_{ij}. All these parameters are assumed to be constant during the option's lifetime $[0, T]$.

Based upon the trend derivatives introduced by [4], and following [7], we consider three types of RTO's and their different payoffs at maturity T, and K representing the strike price, as usual

- *simple* RTO:

$$V^S(T) = \max(\hat{S}_{m'}(T) - K, 0), \quad m' = \arg\max \hat{S}_i(T), \quad 1 \leq i \leq m, \tag{7}$$

- *pure* RTO:

$$V^P(T) = \max(\hat{S}_{m'}(T) - S_{m'}(T), 0), \quad m' = \arg\max \hat{S}_i(T), \quad 1 \leq i \leq m, \tag{8}$$

- *Asian* RTO:

$$V^A(T) = \max(\hat{S}_{m'}(T) - \bar{S}_{m'}(T), 0), \quad m' = \arg\max \hat{S}_i(T), \quad 1 \leq i \leq m, \tag{9}$$

here $\hat{S}_{m'}(T)$ and $\bar{S}_{m'}(T)$ represent the trend and geometric average variables, respectively, of the i-th underlying asset over the option life, $[0, T]$.

Following [7], we may describe the pricing framework that can evaluate several types of RTOs in following way. The option life, $[0, T]$, is partitioned into n time intervals, each of length $\Delta t = \frac{T}{n}$. Hence, the sampling dates for the trend/average options include $t_0 = 0, t_1 = \Delta t, \ldots, t_n = n\Delta t = T$.

According to [4], the trend variable of the i-th asset at maturity can be obtained by solving the following exponential regression model

$$\frac{S_i(t_k)}{S_i(t_0)} = \exp(B_i(t_k - t_0) + \epsilon), \quad k = 0, \ldots, n, \tag{10}$$

where ϵ is a standard white noise. Next, the ordinary least squartes method is employed to solve the estimator

$$\hat{B}_i = \arg\min \sum_{k=0}^{n} \big(\ln \frac{S_i(t_k)}{S_i(t_0)} - B_i(t_k - t_0)\big)^2, \quad B_i \in \mathbb{R} \tag{11}$$

Equipped with $\hat{B}_i$, the exponential trend at maturity can be expressed as

$$\hat{S}_i(T) = S_i(t_0)\exp\hat{B}_i(T-t_0) = S_i(t_0)\exp\big(\sum_{k=1}^{n} c_k \ln\frac{S_i(t_k)}{S_i(t_{k-1})}\big), \tag{12}$$

where

$$c_k = (T-t_0)\sum_{h=k}^{n} b_h = \frac{12\sum_{h=k}^{n}(k\Delta t - \frac{1}{2}T)}{n(n+1)(n+2)(\delta t)^2}T = \frac{6k(n-k+1)}{(n+1)(n+2)}, \quad ,k=1,\dots,n. \tag{13}$$

For pricing different RTOs and keeping line with [7], we consider a general payoff function $V(T)$ containing two new terms $\mathcal{X}(T)$ and $\mathcal{Y}(T)$, defined in following way

$$\begin{gathered} V(T) = \max(\mathcal{X}(T) - \kappa\mathcal{Y}(T), 0), \\ \mathcal{X}(T) = X_{m'}(T),\ \mathcal{Y}(T) = Y_{m'}(T), \quad m' = \arg\max X_i(T), \quad i=1,\dots,m, \\ X_i(T) = \exp\big(\sum_{k=0}^{n}\alpha_k R_{i,k}\big), \quad Y_i(T) = \exp\big(\sum_{k=0}^{n}\beta_k R_{i,k}\big), \quad i=1,\dots,m, \end{gathered} \tag{14}$$

where κ is constant, and $\{\alpha_k\}, \{\beta_k\}$, $k=0,\dots,n$ are two sets of real coefficients unless all being zero simultaneously. The variables $R_{i,k}$ are defined as follows

$$R_{i,0} = \ln S_i(t_0),\ R_{i,k} = \ln\big(\frac{S_i(t_k)}{S_{k-1}}\big), \quad ,k=1,\dots,n$$

where $R_{i,k}, k=1,\dots,n$ represent logarithm returns of the i-th asset for the interval $(t_{k-1}, t_k]$ which follow normal distribution with mean and variance both depending upon Δt

$$R_{i,k} \sim N(\mu_i\Delta t, \sigma_i^2\Delta t),\ \mu_i = r - q_i - \frac{1}{2}\sigma_i^2, \quad k=1,\dots,n.$$

Using definitions of $X_i(T), Y_i(T)$, and distributional properties of $R_{i,k}, k=1,\dots,n$, one gets distributions of the logarithms $\ln X_i(T), \ln Y_i(T)$ which depend upon selected time $t_h = h\Delta t, h=0,\dots,n$, parametrically

$$\begin{gathered} \ln X_i(T) \sim N(\mu_{X_i}, \sigma_{X_i}^2), \qquad \sigma_{X_i}^2 = \sigma_i^2\Delta t\sum_{k=h+1}^{n}\alpha_h^2, \\ \mu_{X_i} = \alpha_0\ln S_i(t_0) + \sum_{k=1}^{h}\alpha_k\ln\frac{S_i(t_k)}{S_i(t_{k-1})} + \mu_i\Delta t\sum_{k=h+1}^{n}\alpha_k. \end{gathered} \tag{15}$$

$$\begin{gathered} \ln Y_i(T) \sim N(\mu_{Y_i}, \sigma_{Y_i}^2), \qquad \sigma_{Y_i}^2 = \sigma_i^2\Delta t\sum_{k=h+1}^{n}\beta_h^2, \\ \mu_{Y_i} = \alpha_0\ln S_i(t_0) + \sum_{k=1}^{h}\beta_k\ln\frac{S_i(t_k)}{S_i(t_{k-1})} + \mu_i\Delta t\sum_{k=h+1}^{n}\beta_k. \end{gathered} \tag{16}$$

For correlations, one gets the following expressions, whereas refering to [7] for more technical details

$$\begin{gathered} \mathrm{Corr}\big(\ln X_i(T), \ln X_j(T)\big) = \rho_{X,ij} = \mathrm{Corr}\big(\ln Y_i(T), \ln Y_j(T)\big) = \rho_{Y,ij} = \rho_{ij}, \\ \mathrm{Corr}\big(\ln X_i(T), \ln Y_i(T)\big) = \rho_{XY,i} = \frac{\sum_{k=h+1}^{n}\alpha_k\beta_k}{\sqrt{(\sum_{k=h+1}^{n}\alpha_k^2)}\sqrt{(\sum_{k=h+1}^{n}\beta_k^2)}}. \end{gathered} \tag{17}$$

Using traditional risk-neutral valuation technique and employing the martingale pricing method, the arbitrage-free price of RTO with the general payoff function, given by expression (14), can be expressed as the expected present value of the payoff at maturity under the risk-neutral measure in following form

$$\begin{gathered} V(t_k) = e^{-r(T-t_k)}E^Q[V(T)] \\ = e^{-r(T-t_k)}E^Q[(X_{m'}(T) - \kappa Y_{m'}(T)) \times I_{\{X_{m'}(T)\geq\kappa Y_{m'}(T)\}} | m' = \arg\max X_i(T),\ i=1,\dots,m] \end{gathered} \tag{18}$$

where $I_{\{.\}}$ is the usual indicator function returning unity when the event specified in the braces is true, otherwise returning zero. Finally, based upon [7] with all details therein, one gets the closed-form formula for pricing RTO

$$\begin{aligned} V(t_k) &= e^{-r(T-t_k)} \sum_{i=1}^{m} \exp(\mu_{X_i} + \frac{1}{2}\sigma_{X_i}^2) N_m\big(\{d_{Y_j,X_i}^{Q_{X_i}}\}_{1\le j\ne i<m}, d_{Y_i,X_i}^{Q_{X_i}}; R^{Q_{X_i}}\big) \\ &-\kappa\, e^{-r(T-t_k)} \sum_{i=1}^{m} \exp(\mu_{Y_i} + \frac{1}{2}\sigma_{Y_i}^2) N_m\big(\{d_{Y_j,X_i}^{Q_{Y_i}}\}_{1\le j\ne i<m}, d_{Y_i,X_i}^{Q_{Y_i}}; R^{Q_{Y_i}}\big) \end{aligned} \tag{19}$$

where $N_m(d_{a,b}^{\pi}; R^{\pi})$ denotes a multivariate normal cumulative distribution function with m parameters $d_{a,b}^{\pi}$ serving as means, and $m \times m$ correlation matrix R^{π}. The indeces a, b, and π stand as short-hands of $a = X_j, Y_i$, $b = X_i$, and $\pi = Q_{X_i}$, Q_{Y_i}, which denote probability measures that are equivalent to the risk-neutral measure Q and facilitates the derivation of the pricing formula.

Note: As far as an analytic structure, the formula(19) resembles the classical Black-Scholes formula for pricing European call option on asset S following the geometric Brownian motion, with volatility σ, exercise price K, q dividend rate, and risk-free interest rate r

$$C(t) = e^{q(T-t)} S N(d_1) - K e^{r(T-t)} N(d_2), \quad d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{(T-t)}},\ d_2 = d_1 - \sigma\sqrt{(T-t)},$$

where $N(.)$ is the cumulative normal distribution.

From numerical point of view, the crucial problem of application pricing formula (19), is an effective calculation of $N_m(d_{a,b}^{\pi}; R^{\pi})$.

The general pricing formula (19) can be used in various ways. The most simple one is to evaluate the simple RTO having the payoff $V^S(T)$, given by Eq.(7), which is actually defined as the difference between the maximum realized trend of m underlying assets $S_i, i = 1, \ldots, m$ and a fixed strike price K. Now, the corresponding exponential trend price at maturity $\hat{S}_i(T)$, expressed in general by Eq.(12), takes the following form

$$\begin{aligned} \hat{S}_i(T) &= S_i(t_0) \exp\big(\sum_{k=1}^{n} c_k \ln \frac{S_i(t_k)}{S_i(t_{k-1})}\big) = \exp\big(\ln S_i(t_0) + \sum_{k=1}^{n} c_k \ln \frac{S_i(t_k)}{S_i(t_{k-1})}\big) \\ &= \exp\big(R_{i,0} + \sum_{k=1}^{n} c_k R_{i,k}\big) \end{aligned} \tag{20}$$

where $\alpha_0 = 1, \alpha_k = c_k = \frac{6k(n-k+1)}{(n+1)(n+2)}, k = 1, \ldots, n$ $\beta_k = 0, k = 0, \ldots, n$, and $\kappa = K$, which provides replacement $X_i(T), \kappa Y_i(T)$ by $\hat{S}_i(T), K$, in Eq.(14), respectively. Hence, the Eq.(14) turns into $V(T) = \max(\hat{S}_{m'}(T) - K, 0)$, $m^{'} = \arg\max \hat{S}_i(T)$, $1 \le i \le m$, which is identical with $V^S(T)$, given by Eq.(7).

Now, using ideas from [7], we can recast the general pricing formula (19) into the pricing formula for the simple RTO, denoted $V^S(t_k)$, and given by Eq.(21), when adopting simpler notation

$$\begin{aligned} & d_j^1 \equiv d_{Y_j,X_i}^{Q_{X_i}},\ d_i^1 \equiv d_{Y_i,X_i}^{Q_{X_i}}, \quad d_j^2 \equiv d_{Y_j,X_i}^{Q_{Y_i}},\ d_i^2 \equiv d_{Y_i,X_i}^{Q_{Y_i}}, \quad \text{and } R^1 \equiv R^{Q_{X_i}},\ R^2 \equiv R^{Q_{Y_i}}, \\ & V^S(t_k) = e^{-r(T-t_k)} \sum_{i=1}^{m} \hat{S}_i(t_k) \exp\big(\mu_i \Delta t \sum_{h=k+1}^{n} c_k + \frac{1}{2}\sigma_{X_i}^2\big) N_m\big(\{d_j^1\}_{1\le j\ne i<m}, d_i^1; R^1\big) \\ & \qquad -K\, e^{-r(T-t_k)} \sum_{i=1}^{m} N_m\big(\{d_j^2\}_{1\le j\ne i<m}, d_i^2; R^2\big), \\ & \qquad \hat{S}_i(t_k) = S_i(t_0) \exp(\sum_{h=1}^{k} c_h \ln \frac{t_h}{t_{h-1}} \end{aligned} \tag{21}$$

where $\hat{S}_i(t_k)$ gives the realized trend price for the i-th asset S_i up to time t_k. The variables appearing in Eq.(21)

are defined as follows

$$d_j^1 = \frac{1}{\sigma_{X,ij}}\big(\ln\frac{\hat{S}_i(t_k)}{\hat{S}_j(t_k)} + (\mu_i-\mu_j)\Delta t\sum_{h=k+1}^{n} c_h + (\sigma_{X_i})^2 - \rho_{X,ij}\sigma_{X_i}\sigma_{X_j}\big),\ \text{for } 1\le j\ne i\le m,$$

$$d_i^1 = \frac{1}{\sigma_{X_i}}\big(\ln\frac{\hat{S}_i(t_k)}{K} + \mu_i\Delta t\sum_{h=k+1}^{n} c_h + (\sigma_{X_i})^2 - \rho_{X,ij}\sigma_{X_i}\sigma_{X_j}\big),$$

$$d_j^2 = \frac{1}{\sigma_{X,ij}}\big(\ln\frac{\hat{S}_i(t_k)}{\hat{S}_j(t_k)} + (\mu_i-\mu_j)\Delta t\sum_{h=k+1}^{n} c_h\big),\ \text{for } 1\le j\ne i\le m,\quad d_i^2 = \frac{1}{\sigma_{X_i}}\big(\ln\frac{\hat{S}_i(t_k)}{K} + \mu_i\Delta t\sum_{h=k+1}^{n} c_h\big),$$

$$R^1 = R^2 = \begin{bmatrix} A_{(m-1)\times(m-1)} & B_{(m-1)\times 1} \\ B^T_{1\times(m-1)} & 1 \end{bmatrix}_{m\times m},\quad B_{(m-1)\times 1} = [\frac{\sigma_{X_i}-\rho_{X,ij}\sigma_{X_j}}{\sigma_{X,ij}}]_{(m-1)\times 1},\ \text{for } 1\le j\ne i\le m,$$

$$A_{(m-1)\times(m-1)} = [\frac{\sigma^2_{X_i}-\rho_{ij}\sigma_{X_i}\sigma_{X_j}-\rho_{il}\sigma_{X_i}\sigma_{X_l}+\rho_{jl}\sigma_{X_j}\sigma_{X_l}}{\sigma_{X,ij}\,\sigma_{X,il}}]_{(m-1)\times(m-1)},\ \text{for } 1\le j,l\ne i\le m.$$

In accordance with Eqs.(15) and (17), $\sigma^2_{X_i} = \sigma_i^2\Delta t\sum_{h=k+1}^{n} c_h^2$, and further $\sigma^2_{X,ij} = \sigma^2_{X_i} - 2\rho_{ij} + \sigma^2_{X_j}$. Here, B^T denotes transposition of any matrix B, as usual.

Input data and *Mathematica functions*:
Input data: $T, K, r, \{q_i\}, \{\sigma_i\}, \{\rho_{ij}\}, i, j = 1,\dots,m,\quad \{S_i(t_k)\}, i = 1,\dots,m,\ j = 1,\dots,n$
Mathematica functions proposed to be used for numerical computations with (19), and (21), in particular:
data ... downladed prices $\{S_i(t)\}, i = 1,\dots,m,\ t\in[0,T]$, or simulated, e.g. td;
td = RandomFunction[WienerProcess[1,1],{0,5,0.05},100] ... to simulate 5 trajectories in [0,5] by step 0.05;
FindFit[data, model, parameters, variable], ... finds numerical values of the parameters that make model give a least-squares fit to data as a function of a variable, e.g. model = a*Exp[-k*t], parameters = {a,k}, variable = t;
MultinormalDistribution[μ,Σ] ... represents a multivariate normal distribution with mean vector μ and covariance matrix Σ;
CDF[MultinormalDistribution[μ,Σ], $\{s_1,\dots,s_m\}$] ... gives the cumulativedistribution function for the multvariate normal distribution evaluated at $\{s_1,\dots,s_m\}$;

4 Conclusion

In the paper, we discuss briefly multi-asset European option pricing problem using PDE approach. Special attention was given to formulation of payoff functions representing terminal conditions, and boundary conditions, as well. On the 2-dimensional example, we also elucidate modifications of boundary conditions due to degeneration. Further, we discuss main ideas of pricing framework of rainbow trend/average options, and simple RTO in particular. From numerical point of view, we mention that efficient calculation of cumulative distribution function of multivariate normal distribution is crucially important attribute therein.

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] Fries, Ch.: *Mathematical finance – theory, modeling, implementation.* John Wiley & Sons Inc., Hoboken, New Jersey, 2007.

[4] Leippold, M., and Syz, J.: Trend derivatives: Pricing, hedging, and application to executive stock options. *Journal of Futures Markets* **27** (2007), 113-129.

[5] Lukáš, L.: Correlation Structure of Underlying Assets Affecting Multi-asset European Option Price. In: *Conf. Proceedings, 35-th Int. Conf. Mathematical Methods in Economics 2017*, (Pražák, P., editor.). Univ. of Hradec Králové, Hradec Králové, 2017, 414–419.

[6] Shiryaev, A.N.: *Essentials of Stochastic Finance – Facts, Models, Theory*. World Scientific Publ. Co., Singapore, 2008.

[7] Wang, J-Y., Wang, H-Ch., Ko, Y-Ch., and Hung, M-W.: Rainbow trend options: valuation and applications. *Rev. Deriv. Res.* **20** (2017), 91-133. DOI 10.1007/s11147-016-9125-z.

Two-assets portfolio with trapezoidal oriented fuzzy present values

Anna Łyczkowska-Hanćkowiak[1], Krzysztof Piasecki[2],

Abstract. The main goal of this paper is to present characteristics of a two-asset portfolio for the case when composing securities are evaluated by imprecise present values modelled by an ordered trapezoidal fuzzy number. The future value is determined as random variable under the Gaussian distribution of probability. The expected discount factor is defined for the case of simple return rate. Obtained model is an extension of the Markowitz theory to the case of securities present values given by ordered fuzzy numbers. Throughout the analysis a fuzzy expected discount factor and imprecision risk assessments are calculated. Thanks to that, there arises a possibility to describe the impact of portfolio diversification on imprecision risk. Presented theoretical results are supported by numerical example.

Keywords: excepted discount factor, ordered fuzzy number, portfolio, present value

JEL Classification: C44, C02, G10
AMS Classification: 03E72, 91G10

1 Introduction

By a financial asset we understand an authorization to receive a future financial revenue, payable to a certain maturity. The value of this revenue is interpreted as anticipated future value (FV) of the asset. In [13, 14] it is detailed justified that if the financial asset is a security then its FV is a random variable. Some security evaluations may be given as function of FV and present value (PV) defined as a current equivalent of a payment available in a given time in the future. PV is widely accepted to be an approximate value, with fuzzy numbers being one of the main tools of its modelling [16, 17]. A detailed description of the evolution of this particular model can be found in [11, 14]. Kuchta [7] has proved the sensibility of trapezoidal fuzzy numbers as a fuzzy financial arithmetic tool. Therefore, the main focus of our research is security with PV described by trapezoidal ordered fuzzy number (OFN) [4].

In [13] the case of portfolio with trapezoidal fuzzy present value was researched. There it is shown that, for appraising the considered securities, the discount factor is better tool than return rate. In [9] discount factor is considered for the case securities with PV given by OFN. The main goal of this paper is to study discount factor of portfolio containing securities with PV described by OFN. The proposed portfolio model will be an extension of the classic Markowitz model to the fuzzy case. The imprecision risk and uncertainty risk are also examined. For a better perception of the reasoning description, in this article we will limit ourselves to considering the case of a two-asset portfolio. By using the mathematical induction, all results obtained this way can be generalized to the case of a multi-asset portfolio. The directions for further research will be indicated.

2 Basic notions and facts

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 as a fuzzy subset of the real line $\mathbb{R}$. The most general definition of FN was given by Dubois and Prade [1]. Moreover, they have introduced [2] such arithmetic operations on FN which are coherent with the Zadeh Extension Principle.

The original Kosiński's concept of ordered fuzzy numbers (OFN) [4] was revised in [12]. In general, an OFN is defined as FN supplemented by negative orientation or positive one. Negative orientation means the order from bigger numbers to smaller ones. Positive orientation means the order from smaller numbers to bigger ones.The space of all OFN is denoted by the symbol $\mathbb{K}$. In this paper, we will limit our considerations to the following kind of OFN.

[1] WSB University in Poznań, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland, e-mail:anna.lyczkowska-hanckowiak@wsb.poznan.pl
[2] Poznań University of Economics, Department of Investment and Real Estate, al. Niepodległości 10, 61-875 Poznań, Poland, E-mail: krzysztof.piasecki@ue.poznan.pl

Definition 1.[12] *For any monotonic sequence* $(a, b, c, d) \subset \mathbb{R}$ *the trapezoidal ordered fuzzy number (TrOFN)* $\overleftrightarrow{Tr}(a, b, c, d)$ *is defined as the pair of FN determined by its membership function* $\mu_{\overleftrightarrow{Tr}}(\cdot\,|a, b, c, d) \in [0; 1]^{\mathbb{R}}$ *given by the identity*

$$\mu_{\overleftrightarrow{Tr}}(x|a, b, c, d) = \begin{cases} 0 & x \notin [a, d] = [d, a] \\ \frac{x-a}{b-a} & x \in [a, b[=]b, a] \\ 1 & x \in [b, c] = [c, b] \\ \frac{x-d}{c-d} & x \in]c, d] = [d, c[\end{cases} \tag{1}$$

and orientation from $a \in \mathbb{R}$ to $b \in \mathbb{R}$. □

The sum $\boxplus$ is commutative. Moreover, due results obtained in [5] and [12], we can say that if OFNs have identical orientation then their K-sum $\oplus$ and sum $\boxplus$ are identical with sum obtained by means of arithmetic introduced by Dubois and Prade [2].

OFNs are widely used as a model for assessment or estimation of a parameter, which is given imprecisely [15]. After [3] we understand imprecision as a superposition of ambiguity and indistinctness of information. Ambiguity can be interpreted as a lack of a clear recommendation between one alternative among various others. Indistinctness is understood as a lack of explicit distinction between recommended and not recommended alternatives. An increase in financial information imprecision makes investors' decisions riskier and therefore it is logical to consider the problem of imprecision assessment. We measure the ambiguity of an arbitrary TrOFN by the energy measure $d: \mathbb{K} \to \mathbb{R}_0^+$ [8]. defined here in following way

$$d\left(\overleftrightarrow{Tr}(a, b, c, d)\right) = \frac{1}{2}|d + c - b - a| \,. \tag{5}$$

The indistinctness of an arbitrary TrOFN can be measured by the entropy measure $e: \mathbb{K} \to \mathbb{R}_0^+$ [6] given here in following form

$$e\left(\overleftrightarrow{Tr}(a, b, c, d)\right) = \frac{d-c+b-a}{3 \cdot d+c-b-3 \cdot a} \,. \tag{6}$$

3 Oriented discount factor

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 and assessed PV V_0. The basic characteristic of benefits from owning this security is a return rate r_t given by the identity

$$r_t = \frac{V_t - V_0}{V_0}. \tag{7}$$

In [13, 14] it is justified 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

$$\tilde{r}_t(\omega) = \frac{\tilde{V}_t(\omega) - \check{C}}{\check{C}} \tag{8}$$

In practice of financial markets analysis, the uncertainty risk is usually described by probability distribution of return rate determined by (8). At the moment, we have an extensive knowledge on this subject. After Markowitz [10] we assume that the return rate $\tilde{r}_t$ has a Gaussian distribution $N(\bar{r}, \sigma)$ of probability.

In this paper, due Kuchta [7] we additionally assume that the PV is imprecisely estimated by oriented PV given as TrOFN $\overleftrightarrow{PV} = \overleftrightarrow{Tr}(V_s, V_f, V_l, V_e)$ where the monotonic sequence $\{V_s, V_f, \check{C}, V_l, V_e\}$ is interpreted in the following way:

– $[V_s, V_e] \subset \mathbb{R}^+$ is interval of all possible PV' values,

– $[V_f, V_l] \subset [V_s, V_e]$ is interval of all prices which do not perceptible differ from market price $\check{C}$.

In [9] it is shown that then considered security is characterized is by oriented discount factor $\mathcal{V} \in \mathbb{K}$ determined with use the formula

$$\overleftrightarrow{\mathcal{V}} = \overleftrightarrow{Tr}\left(\frac{V_s}{\check{C}} \cdot \bar{v}, \frac{V_f}{\check{C}} \cdot \bar{v}, \frac{V_l}{\check{C}} \cdot \bar{v}, \frac{\check{C}_e}{\check{C}} \cdot \bar{v}\right), \tag{9}$$

where expected discount factor $\bar{v}$ is determined as follows

$$\bar{v} = \frac{1}{1+\bar{r}}. \tag{10}$$

The formal simplicity of obtained description of an expected discount factor encourages for its further application as a portfolio analysis tool. The maximization criterion of expected return rate can then be substituted by minimization criterion of the expected discount factor. In case of non-fuzzy values of both parameters, those criteria are equivalent.

An increase in ambiguity of expected discount factor $\mathcal{V} \in \mathbb{F}$ suggests a higher number of alternative recommendations to choose from. This may result in making a decision, which will be *ex post* associated with profit lower than maximal, so with lost chances. This kind of risk is called an ambiguity risk. The ambiguity risk implied by $\mathcal{V}$ is measured by energy measure $d(\mathcal{V})$.

An increase in the indistinctness of $\mathcal{V}$, on the other hand, suggests that the differences between recommended and un-recommended decision alternatives are harder to differentiate. This leads to an increase in the indistinctness risk, which is in a risk of choosing a not recommended option. The indistinctness risk of an expected discount factor $\mathcal{V}$ is measured by entropy measure $e(\mathcal{V})$. Imprecision risk consists of both ambiguity and indistinctness risk combined.

From (14) we have, that the return rate is a function of the future value of an asset, which is uncertain, since we don't know the future state of the world. Because of this, the investor is not sure, whether they will gain or lose from the decision they made. With the increase in uncertainty, the risk of making a wrong decision is higher. In this paper, behind Markowitz [10] uncertainty risk of a return rate will be measured by its variance σ^2 .

4 Two-assets portfolio

Let us consider the case of a two-asset portfolio π, consisting of financial assets Y_1 and Y_2. The oriented PV of assets Y_i $(i = 1; 2)$ is estimated by fuzzy TrOFN $\overleftarrow{PV}^{(i)} = \overleftarrow{Tr}\left(V_s^{(i)}, V_f^{(i)}, V_l^{(i)}, V_e^{(i)}\right)$. We assume that for each security $Y_i (i = 1; 2)$ we know the simple return rate $\tilde{r}_t^i: \Omega \to \mathbb{R}$ appointed for market price $\check{C}^{(i)}$ by (8). After Markowitz [10] we assume that the two-dimensional variable $(\tilde{r}_t^1, \tilde{r}_t^2)^T$ has a cumulative Gaussian distribution $N((\bar{r}_1, \bar{r}_2)^{\mathrm{T}}, \mathbf{\Sigma})$, with a covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & cov_{12} \\ cov_{12} & \sigma_2^2 \end{pmatrix}. \tag{11}$$

We appoint an oriented discount factor of security Y_i:

$$\overleftarrow{\mathcal{V}}^{(i)} = Tr\left(V_s^{(i)} \cdot \frac{\bar{v}_i}{\check{C}^{(i)}}; V_f^{(i)} \cdot \frac{\bar{v}_i}{\check{C}^{(i)}}; V_l^{(i)} \cdot \frac{\bar{v}_i}{\check{C}^{(i)}}; V_e^{(i)} \cdot \frac{\bar{v}_i}{\check{C}^{(i)}}\right) = \frac{\bar{v}_i}{\check{C}^{(i)}} \odot \overleftarrow{PV}^{(i)}, \tag{12}$$

where $\bar{v}_i$ is an expected discount factor appointed using the expected return rate $\bar{r}_i$. We have that the market value $\check{C}^{(\pi)}$ of a portfolio π is equal to

$$\check{C}^{(\pi)} = \check{C}^{(1)} + \check{C}^{(2)}. \tag{13}$$

Share p_i of an instrument Y_i in the portfolio π is given by

$$p_i = \frac{\check{C}^{(i)}}{\check{C}^{(\pi)}}. \tag{14}$$

Then portfolio expected discount factor $\bar{v}_\pi$ equals[13]

$$\bar{v}_\pi = \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1}. \tag{15}$$

The portfolio variance σ_π^2 can be calculated by the formula

$$\sigma_\pi^2 = p_1^2 \cdot \sigma_1^2 + 2 \cdot p_1 \cdot p_2 \cdot cov_{12} + p_2^2 \cdot \sigma_2^2 . \tag{16}$$

The portfolio PV is given by the sum

$$\overleftarrow{PV}^{(\pi)} = \overleftarrow{PV}^{(1)} \boxplus \overleftarrow{PV}^{(2)}. \tag{17}$$

In agree with general properties of the sum $\boxplus$, if $\overleftarrow{PV}^{(1)}$ and $\overleftarrow{PV}^{(2)}$ have identical orientation then the portfolio PV $\overleftarrow{PV}^{(\pi)}$ is equal to sum obtained by means of arithmetic introduced by Dubois and Prade [2]. In [13] it is proved that for this case the portfolio oriented discount factor $\mathcal{V}^{(\pi)}$ may be calculated in following way

$$\overleftarrow{\mathcal{V}}^{(\pi)} = \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1} \odot \left(\left(\frac{p_1}{\bar{v}_1} \odot \overleftarrow{\mathcal{V}}^{(1)}\right) \oplus \left(\frac{p_2}{\bar{v}_2} \odot \overleftarrow{\mathcal{V}}^{(2)}\right)\right). \tag{18}$$

Let us consider the case when $\overleftarrow{PV}^{(1)}$ and $\overleftarrow{PV}^{(2)}$ have different orientation. Then

$$\overleftarrow{PV}_{\pi} = \overleftarrow{Tr}\left(V_s^{(\pi)}, V_f^{(\pi)}, V_l^{(\pi)}, V_e^{(\pi)}\right), \tag{19}$$

where

$$S = V_s^{(1)} + V_s^{(2)}, V_f^{(\pi)} = V_f^{(1)} + V_f^{(2)}, V_l^{(\pi)} = V_l^{(1)} + V_l^{(2)}, E = V_e^{(1)} + V_e^{(2)},$$

$$V_s^{(\mu)} = \begin{cases} \min\left\{S, V_f^{(\pi)}\right\} & \left(V_f^{(\pi)} < V_l^{(\pi)}\right) \vee \left(V_f^{(\pi)} = V_l^{(\pi)} \wedge S \leq E\right) \\ \max\left\{S, V_f^{(\pi)}\right\} & \left(V_f^{(\pi)} > V_l^{(\pi)}\right) \vee \left(V_f^{(\pi)} = V_l^{(\pi)} \wedge S > E\right) \end{cases},$$

$$V_e^{(\mu)} = \begin{cases} \max\left\{E, V_l^{(\pi)}\right\} & \left(V_f^{(\pi)} < V_l^{(\pi)}\right) \vee \left(V_f^{(\pi)} = V_l^{(\pi)} \wedge S \leq E\right) \\ \min\left\{E, V_l^{(\pi)}\right\} & \left(V_f^{(\pi)} > V_l^{(\pi)}\right) \vee \left(V_f^{(\pi)} = V_l^{(\pi)} \wedge S > E\right) \end{cases}.$$

This together with (9) implies that for $\overleftarrow{PV}^{(1)}$ and $\overleftarrow{PV}^{(2)}$ having different orientation, the portfolio oriented discount factor equals

$$\vec{\mathcal{V}}^{(\pi)} = \overleftarrow{Tr}\left(V_s^{(\pi)} \cdot \frac{\bar{v}}{\check{C}^{(\pi)}}; V_f^{(\pi)} \cdot \frac{\bar{v}}{\check{C}^{(\pi)}}; V_l^{(\pi)} \cdot \frac{\bar{v}}{\check{C}^{(\pi)}}; V_e^{(\pi)} \cdot \frac{\bar{v}}{\check{C}^{(\pi)}}\right) = \frac{\bar{v}}{\check{C}^{(\pi)}} \odot \overleftarrow{PV}^{(\pi)}. \tag{20}$$

In the next section we will study among other things such case of $\overleftarrow{PV}^{(1)}$ and $\overleftarrow{PV}^{(2)}$ with different orientation, when the discount factor (19) does not satisfy the condition (17).

5 Numerical example

We will consider portfolios containing only two securities: $\pi = \{Y_1, Y_2\}$ and $\varpi = \{Y_1, Y_3\}$. For each security Y_i $(i = 1,2,3)$ information about its expected return rate $\bar{r}_i$, variance σ_i^2 of its return rate, covariance $cov_{1,i}$, its market price $\check{C}^{(i)}$ and its oriented PV $\overleftarrow{PV}^{(i)}$ are summarized in Table 1.

	Securities		
Parameters	$\boldsymbol{Y_1}$	$\boldsymbol{Y_2}$	$\boldsymbol{Y_3}$
$\bar{r}_i$	0.25	0.50	0.50
σ_i^2	0.5	0.4	0.4
$cov_{1,i}$		-0.1	-0.1
$\check{C}^{(i)}$	24	69	69
$\overleftarrow{PV}^{(i)}$	$\overleftarrow{Tr}(18,23,25,27)$	$\overleftarrow{Tr}(66,67,70,75)$	$\overleftarrow{Tr}(76,72,69,67)$

Table 1. The basic information on considered securities

In the next step, we calculate for each security Y_i $(i = 1,2,3)$ in turn:
- using (10) its expected discount factor $\bar{v}_i$,
- using (14) its share p_i in portfolio,
- using (12) its oriented discount factor $\vec{\mathcal{V}}^{(i)}$,
- using (5) energy measure $d\left(\vec{\mathcal{V}}^{(i)}\right)$,
- using (6) entropy measure $e\left(\vec{\mathcal{V}}^{(i)}\right)$.

Obtained evaluations are summarized in the Table 2. We see that the securities from the portfolio ϖ differ in PV orientation. The securities from the portfolio π have identical PV orientation. The fact that we take into account when evaluating considered portfolios. We calculate for each portfolio $\varrho \in \{\pi, \varpi\}$ in turn:
- using (15) its expected discount factor $\bar{v}_{\varrho}$,
- using (16) the variance σ_{ϱ}^2 of its return rate,
- using (18) oriented discount factor $\vec{\mathcal{V}}^{(\pi)}$,

- using (19) oriented discount factor $\overleftarrow{\mathcal{V}}^{(\varpi)}$,
- using (5) energy measure $d\left(\overleftarrow{\mathcal{V}}^{(\varrho)}\right)$,
- using (6) entropy measure $e\left(\overleftarrow{\mathcal{V}}^{(\varrho)}\right)$.

Obtained evaluations are summarized in the Table 3.

	Securities		
Parameters	$\boldsymbol{Y_1}$	$\boldsymbol{Y_2}$	$\boldsymbol{Y_3}$
$\overline{v}_i$	0.800	0.667	0.667
p_i	0.258	0.742	0.742
$\overleftarrow{\mathcal{V}}^{(i)}$	$\overleftarrow{Tr}(0.600, 0.767, 0.833, 0.900)$	$\overleftarrow{Tr}(0.638, 0.648, 0.677; 0.725)$	$\overleftarrow{Tr}(0.735, 0.696, 0.667, 0.648)$
$d\left(\overleftarrow{\mathcal{V}}^{(i)}\right)$	0.183	0.058	0.058
$e\left(\overleftarrow{\mathcal{V}}^{(i)}\right)$	0.242	0.200	0.200

Table 2. Evaluations of considered securities

	Portfolios	
Parameters	$\boldsymbol{\pi = \{Y_1, Y_2\}}$	$\boldsymbol{\varpi = \{Y_1, Y_3\}}$
$\overline{v}_\varrho$	0.697	0.697
σ_ϱ^2	0.22	0.22
$\overleftarrow{\mathcal{V}}^{(\varrho)}$	$\overleftarrow{Tr}(0.629, 0.675, 0.712, 0.764)$	$\overleftarrow{Tr}(0.712, 0.712, 0.704, 0.704)$
$d\left(\overleftarrow{\mathcal{V}}^{(\varrho)}\right)$	0.086	0.008
$e\left(\overleftarrow{\mathcal{V}}^{(\varrho)}\right)$	0.220	0.000

Table 3. Evaluations of considered portfolios

We obtain following relations between the energy measure and entropy measure appointed for fuzzy expected discount factors of portfolio and its components:

$$d\left(\overleftarrow{\mathcal{V}}^{(1)}\right) > d\left(\overleftarrow{\mathcal{V}}^{(\pi)}\right) > d\left(\overleftarrow{\mathcal{V}}^{(2)}\right) = d\left(\overleftarrow{\mathcal{V}}^{(3)}\right) > d\left(\overleftarrow{\mathcal{V}}^{(\varpi)}\right),$$

$$e\left(\overleftarrow{\mathcal{V}}^{(1)}\right) > e\left(\overleftarrow{\mathcal{V}}^{(\pi)}\right) > e\left(\overleftarrow{\mathcal{V}}^{(2)}\right) = e\left(\overleftarrow{\mathcal{V}}^{(3)}\right) > e\left(\overleftarrow{\mathcal{V}}^{(\varpi)}\right).$$

These inequalities show that for the case of securities with identical PV orientation the portfolio diversification can average the imprecision risk. Moreover, we see that for the case of securities with different PV orientation the portfolio diversification can reduce the imprecision risk. Additionally, we have

$$\sigma_1^2 > \sigma_2^2 = \sigma_3^2 > \sigma_\pi^2 = \sigma_\varpi^2.$$

It shows that uncertainty risk and imprecision risk may be reduced in the same time.

Let's try to determine the oriented discount factor $\overleftarrow{\mathcal{V}}^{(\varpi)}$ using identity (18)

$$\mathcal{V}^{(\varpi)} = \left(\left(\left(\frac{0.258}{0.8} + \frac{0.742}{0.67}\right)^{-1} \cdot \frac{0.258}{0.8}\right) \odot \mathcal{V}^{(1)}\right) \oplus \left(\left(\left(\frac{0.258}{0.8} + \frac{0.742}{0.67}\right)^{-1} \cdot \frac{0.742}{0.67}\right) \odot \mathcal{V}^{(3)}\right) =$$

$$= \left(0.2257 \odot \overleftarrow{Tr}(0.600, 0.767, 0.833, 0.900)\right) \oplus \left(0.7743 \odot \overleftarrow{Tr}(0.712; 0.712; 0.704; 0.704)\right) =$$

$$= Tr(0.706, 0.712, 0.704, 0.705)$$

The sequence $\{0.706, 0.712, 0.704, 0.705\}$ is not monotonic. It implies that above K-sum does not exist. It means that the oriented discount factor portfolio $\overleftarrow{\mathcal{V}}^{(\varpi)}$ cannot be calculated by analogous way as the oriented discount factor portfolio $\overleftarrow{\mathcal{V}}^{(\pi)}$ by the identity (18).

6 Summary

The main purpose of this article was to analyze the possibility of managing the risk burdening a two-asset portfolio, built with use of imprecise information stemming from PV of component assets. The imprecise PVs were modelled with by TrOFN. For this assumption we have reached the following conclusions:

- the portfolio diversification along with diversification of PV orientation can lower imprecision risk;
- the portfolio diversification without diversification of PV orientation can simultaneously lower uncertainty risk and imprecision risk at the same time,
- the portfolio diversification can simultaneously reduce that uncertainty risk and imprecision risk.

Obtained results suggest that the PV orientation may useful tools for managing interaction of imprecision risk and uncertainty risk. Moreover, research suggests that there exist portfolios, which imprecision risk will not be minimized with portfolio diversification, and thus it is vital to create a new risk minimization problem, including all of the risk types. The results obtained above encourage for their broader analysis. Further research can focus on generalizing the representation of the present value to an arbitrary fuzzy number. By using the mathematical induction, all results obtained this way can be generalized to the case of a multi assets portfolio.

References

[1] Dubois D., Prade H. (1978) Operations on fuzzy numbers, *International Journal of System Science* 9, 613-629, doi:10.1080/00207727808941724

[2] Dubois D., Prade H. (1979), Fuzzy real algebra: some results, *Fuzzy Sets and Systems* 2, 327-348

[3] Klir G.J. (1993), Developments in uncertainty-based information, *Advances in Computers* 36, 255-332, doi:10.1016/s0065-2458(08)60273-9

[4] Kosiński W. Prokopowicz, P. Ślęzak D. (2002) Fuzzy numbers with algebraic operations: algorithmic approach. In M. Klopotek M., S.T.Wierzchoń, M. Michalewicz M., (Eds.) *Proc.IIS'2002*,(pp 311-320) Sopot, Poland,: Physica Verlag, Heidelberg .

[5] Kosiński W. (2006) On fuzzy number calculus. *Int. J. Appl. Math. Comput. Sci*. 16(1) 51–57.

[6] Kosko B. (1986), Fuzzy entropy and conditioning, *Information Sciences* 40, 165-174, doi:1016/0020-0255(86)90006-x

[7] Kuchta D., (2000), Fuzzy capital budgeting, *Fuzzy Sets and Systems* 111, 367–385

[8] de Luca A., Termini S., (1979), Entropy and energy measures of fuzzy sets, In M.M. Gupta, R.K. Ragade, R.R. Yager [Eds.], *Advances in fuzzy set theory and applications*, (pp 321-338). New York

[9] Łyczkowska-Hanćkowiak A., Piasecki K., (2018),The expected discount factor determined for present value given as ordered fuzzy number, In W. Szkutnik et al (Eds.) *9th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Winter Edition" Conference Proceedings*, (pp 69-65), Publishing House of the University of Economics in Katowice.

[10] Markowitz H.S.M. (1952), Portfolio Selection, *Journal of Finance* 7, 77-91, doi:10.1111/j.1540-6261.1952.tb01525.x

[11] Piasecki K. (2014), On Imprecise Investment Recommendations, *Studies in Logic, Grammar and Rhetoric*, 37, 179-194, https://doi.org/10.2478/slrg-2014-0024

[12] Piasecki K., (2018) Revision of the Kosiński's theory of ordered fuzzy numbers, *Axioms* 7(1), 16, doi: 10.3390/axioms7010016

[13] Piasecki K., Siwek J., (2018) Two-assets portfolio with trapezoidal fuzzy present values, In W. Szkutnik (Eds.) *9th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Winter Edition" Conference Proceedings*, (pp 76-84) Wyd. UE, Katowice

[14] Piasecki K., Siwek J., (2018), Two-Asset Portfolio with Triangular Fuzzy Present Values—An Alternative Approach, In T. Choudhry, J. Mizerka (Eds.*), Contemporary Trends in Accounting, Finance and Financial Institutions. Proceedings from the International Conference on Accounting, Finance and Financial Institutions (ICAFFI), Poznan 2016*, Springer Proceedings in Business and Economics (pp 11-26) Springer International Publishing, doi:10.1007/978-3-319-72862-9_2

[15] Prokopowicz, P. et al, (Eds.), (2017) *Theory and Applications of Ordered Fuzzy Number. Tribute to Professor Witold Kosiński*, Studies in Fuzziness and Soft Computing, Vol. 356, Springer Verlag, Berlin.

[16] Zadeh L. A., Kacprzyk J. (Eds.), (1999), *Computing with Words in Information/Intelligent Systems 1: Foundations*, Studies in Fuzziness and Soft Computing, Vol. 33, Physica-Verlag, New York.

[17] Zadeh L. A., Kacprzyk J. (Eds.), (1999), *Computing with Words in Information/Intelligent Systems 2: Applications*, Studies in Fuzziness and Soft Computing, Vol. 34, Physica-Verlag, New York.

GARCH models with NIG innovations for stock market indices

Jiří Málek[1], Quang Van Tran[2]

Abstract. GARCH models have been widely used for modeling volatility in various financial engineering applications. In the model the underlying assumption for the distribution of financial asset returns often was a normal distribution. As the distribution of returns substantially differ from the normal distribution as they tend to have heavier tails, some alternative distributions have been investigated. Recently the use of distributions from the generalized hyperbolic distribution family has begun pervading into financial modeling, it is worth exploring how they can be used for capturing the distribution of financial returns in GARCH family models. The aim of this paper is to investigate the suitability of the use of Normal Inverse Gaussian (NIG) distribution, which is a special case of the generalized hyperbolic distribution, in GARCH models. Our study will be performed on stock market indices and the performance of NIG distribution in GARCH models will be compared with those of other distribution alternatives.

Keywords: GARCH models, NIG distribution, Stock market indices, Performance comparison

JEL classification: C13, C58

1 Introduction

The stylized facts on financial asset returns as leptocurticity and volatility clustering have been well documented in the literature. To address this problem, generalized error distribution and Student t-distribution are often used as substitutes for the less suitable normal distribution for modeling volatility clustering with GARCH models. Since the introduction of generalized hyperbolic distribution by Barndorff-Nielsen in 1977 [2], several experiments using the normal inverse gaussian (NIG) distribution as an alternative to the normal one for GARCH model class can be observed [3, 1, 10, 6]. However, due to the complexity of the density of this distribution, its application is not wide-spread, inaccuracies often occur [11, 12, 8]. Moreover, there is a lack of assessment how the use of the NIG distribution can improve the performance of a GARCH model. To fill this gap in the literature, in this paper we will employ the NIG distribution for two most often used models GARCH and GJR-GARCH for volatility clustering modeling of daily returns of four stock market indices: S&P 500 (USA), FTSE 100 (Great Britain), ATX (Austria), and PX (Czech Republic). To implement our research objective, first the parameters of these models are estimated from data using maximum likelihood estimation method. Then we compare the performance of these models with the ones of those using normal distribution and Student t- distribution for the same datasets.

2 Generalized hyperbolic distribution group

The generalized hyperbolic distribution (GH) was first developed by Barndorff-Nielsen [2]. It is one of the alternative distributions for modeling heavy tail property of financial asset returns. It is characterized by five parameters $\theta = (\lambda, \alpha, \beta, \delta, \mu)$, where $\lambda \in \mathbb{R}$, $\alpha > 0$, and $|\beta| \leq \alpha$ are shape parameters, $\delta > 0$ is the scale parameter and μ is the location parameter. Its probability density function is

$$f_{GH}(x) = \kappa[\delta^2 + (x-\mu)^2]^{\frac{1}{2}\left(\lambda-\frac{1}{2}\right)} K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)\exp(\beta(x-\mu)), \tag{1}$$

where

$$\kappa = \frac{(\alpha^2-\beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi}\alpha^{\lambda-\frac{1}{2}}\delta^{\lambda}K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}$$

[1]University of Economics Prague, Department of Probability and Statistics, nám. W. Churchilla 4, Praha 3 - Czech Republic, malek@vse.cz

[2]University of Economics Prague, Department of Probability and Statistics, nám. W. Churchilla 4, Praha 3 - Czech Republic, tran@vse.cz

and K_λ is the modified Bessel function of the third kind with index λ.

All moments of a random variable generalized hyperbolic distributed exist and the first two of them are

$$\mathbb{E}(X) = \mu + \frac{\beta\delta^2}{\zeta}\frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)} \tag{2}$$

$$\mathrm{Var}(X) = \delta^2\left\{\frac{K_{\lambda+1}(\zeta)}{K_\lambda(\zeta)} + \left(\frac{\beta\delta}{\zeta}\right)^2\left[\frac{K_{\lambda+2}(\zeta)}{K_\lambda(\zeta)} - \left(\frac{K_{\lambda+1}(\zeta)}{\zeta K_\lambda(\zeta)}\right)^2\right]\right\} \tag{3}$$

where $\zeta = \delta\sqrt{\alpha^2-\beta^2}$,$\gamma = \sqrt{\alpha^2-\beta^2}$.

The heavy tail property of the generalized hyperbolic distribution comes from the fact that

$$\mathrm{Prob}(X \le x) \approx \mid x \mid^{\lambda-1} \exp((\alpha-\beta)x) \quad \text{as} \quad x \to -\infty. \tag{4}$$

When $\lambda = -\frac{1}{2}$, we get the first special case called the Normal Inverse Gaussian (NIG) distribution and its density becomes

$$f_{\mathrm{NIG}}(x) = \frac{\alpha\delta}{\pi}\exp[\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)]\frac{K_1(\alpha\sqrt{\delta^2+(x-\mu)^2})}{\sqrt{\delta^2+(x-\mu)^2}}. \tag{5}$$

Their corresponding mean and variance are:

$$\mathbb{E}(X) = \mu + \frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}} \tag{6}$$

$$\mathrm{Var}(X) = \frac{\delta\alpha^2}{\left[\sqrt{\alpha^2-\beta^2}\right]^3}. \tag{7}$$

This distribution is often used for modeling financial asset returns instead of the general case above as the parameter estimation of the generalized hyperbolic distribution is accompanied with a severe multimodal and flat objective function problem. Adding another parameter into the model therefore may not substantially increase the goodness of fit of the model, hence NIG distribution seems to be a good approximation for the general case.

3 GARCH models with NIG distribution

Another characteristic feature of financial asset returns is the so called volatility clustering property. This property appears in the form that large changes tend to be followed by large changes of either sign and small changes tend to be followed by small changes. Engle [5] proposes to model this property by an autoregressive conditional heteroskedasticity (ARCH) model. Later Bollerslev [4] generalizes Engle's ARCH model into the GARCH model. A general GARCH model can be specified as

$$r_t = \mu_t(\theta_\mu) + \varepsilon_t \tag{8}$$

$$\sigma_t^2 = g(\sigma_{t-s}^2, \varepsilon_{t-s}; \theta_\sigma) \quad \text{for} \quad s = 1,...,p \tag{9}$$

$$\varepsilon_t \sim F(\theta_F), \tag{10}$$

where μ_t is the conditional mean dependent on a vector of parameters θ_μ, σ_t^2 is the conditional variance dependent on a vector of parameters θ_σ and the residuals ε_t follow a distribution F with a vector of parameters θ_F. When returns are NIG distributed, i. e. $r_t \sim \mathrm{NIG}(\alpha,\beta,\delta,\mu)$, then their first two moments can be described as follows

$$\mathbb{E}(r_t) = \mu + \frac{\beta\delta}{\gamma} \tag{11}$$

$$\mathrm{Var}(r_t) = \frac{\delta\alpha^2}{\gamma^3}, \tag{12}$$

where $\gamma = \sqrt{\alpha^2-\beta^2}$. Let σ_t^2 be the conditional variance of r_t, according to (11) and (12) the conditional mean and parameter δ as well as the mean of r_t can be expressed as

$$\delta = \frac{\gamma^3}{\alpha^2}\sigma_t^2, \tag{13}$$

$$\mathbb{E}(r_t) = \mu + \frac{\beta\gamma^2}{\alpha^2}\sigma_t^2 \tag{14}$$

and the distribution of r_t can be reparameterized as $r_t \sim \text{NIG}\left(\alpha, \beta, \frac{\gamma^3}{\alpha^2}\sigma_t^2, \mu\right)$. From (14) it is clear that the mean of r_t depends on σ_t^2, hence it is the conditional mean, therefore a GARCH model with NIG distribution is also a GARCH-M model. The conditional variance equation of GARCH(1,1) model proposed by Bollerslev [4] with NIG innovations then has the following specification:

$$\sigma_t^2 = a_0 + a_1\sigma_{t-1}^2 + b_1\varepsilon_{t-1}^2 \tag{15}$$

The conditional variance equation of the GJR-GARCH(1,1) model suggested by Glosten et al. [7] is:

$$\sigma_t^2 = a_0 + a_1\sigma_{t-1}^2 + b_1\varepsilon_{t-1}^2 + c_1\varepsilon_{t-1}^2 I_{t-1}, \tag{16}$$

where I_{t-1} is the indicator function and $I_{t-1} = 0$ if $\varepsilon_{t-1} \geq 0$ and $I_{t-1} = 1$ otherwise. This modification of the original GARCH model accounts for the fact that positive and negative unexpected returns may have substantially different impacts on the conditional variance, which is called the leverage effect in the literature.

4 Empirical Analysis and Results

For our empirical analysis, four series of stock market indices S&P 500 (USA), FTSE 100 (Great Britain), ATX (Austria) and PX (Czech Republic) are used. They are daily series of close values of these indices from 2000 to 2018 obtained from Bloomberg database. The original series are converted into logarithmic daily return series. The descriptive statistics of both original series and the transformed series are shown in Table 1.

	Original series				Log-returns series			
	SP500	FTSE	ATX	PX	SP500	FTSE	ATX	PX
mean	1446.3	5737.8	2431.2	985.14	1.36e-4	9.61e-6	2.31e-4	1.82e-4
median	1318.3	5861.9	2388.9	979.42	5.24e-4	3.78e-4	7.38e-4	6.05e-4
mode	1092.5	4489.7	1127.8	407.70	-9.47e-2	-9.27e-2	-1.03e-1	-1.27e-2
min	676.53	3287.0	1003.7	320.10	-9.47e-2	-9.27e-2	-0.103	-0.162
max	2872.9	7788.6	4981.9	1936.1	0.110	9.38e-2	0.120	0.124
std	443.91	949.87	970.00	361.46	1.21e-2	1.19e-2	1.43e-2	1.38e-2
skew	0.955	-0.284	0.588	0.328	-0.219	-0.159	-0.334	-0.478
kurt	3.13	2.35	2.81	2.78	11.61	9.27	10.06	16.01
obs	4735	4735	4735	4554	4561	4582	4493	4538

Table 1 Descriptive statistics of original and logarithmic transformed series

	S&P 500			FTSE 100		
	coeff	S.E.	z-score	coeff	S.E.	z-score
α	131.12	9.700	13.516	190.22	14.11	13.483
β	-3.097	1.400	-2.213	-1.826	1.129	-1.609
μ	5.68e-4	8.31e-4	0.684	2.60e-4	1.19e-4	2.177
a_0	1.35e-6	3.26e-7	4.145	1.88e-6	3.27e-7	5.746
a_1	0.908	9.90e-3	91.787	0.0882	1.06e-2	83.274
b_1	0.0815	9.16e-3	8.899	0.1031	9.87e-3	10.485

Table 2 GARCH(1,1) model estimation results for return series of S&P 500 and FTSE

First, the return series of four stock market indices are used to estimate parameters of GARCH(1,1) and GJR-GARCH(1,1) models. This specification for the (GJR)-GARCH model is based on the work of Hansen and Lunde [9]. Besides parameters of GARCH and GJR-GARCH models, parameters α, β, μ of the distribution of the returns of stock market indices are also estimated at the same time. The maximum likelihood estimation method is applied for the estimation procedure. The estimation is performed in Matlab with our own programming effort as estimation procedure for GARCH models with the NIG distribution are not available in any econometric package yet. The estimation results for model GARCH(1,1) are reported in Tables 2 and 3. For model GJR-GARCH(1,1) they are displayed in Tables 4 and 5. Aside from values of model estimates, their asymptotic standard errors and the corresponding z-score are also computed and shown in these tables [3].

	ATX			PX		
	coeff	S.E.	z-score	coeff	S.E.	z-score
α	124.11	9.742	12.740	155.22	13.074	11.872
β	-3.521	0.631	-5.579	-3.932	0.983	-4.001
μ	8.69e-4	2.52e-4	3.450	9.13e-4	0.400	2.28e-3
a_0	2.88e-6	5.37e-7	5.375	3.12e-6	5.52e-6	0.565
a_1	0.909	1.09e-2	83.649	0.877	1.30e-2	67.618
b_1	0.072	9.46e-3	7.624	0.105	2.30e-2	4.569

Table 3 GARCH(1,1) model estimation results for return series of ATX and PX

	S&P 500			FTSE 100		
	coeff	S.E.	z-score	coeff	S.E.	z-score
α	142.33	4.905	29.018	199.67	2.833	70.586
β	-3.426	0.198	-17.265	-2.031	8.95e-2	-22.677
μ	5.59e-4	1.63e-4	3.429	2.51e-4	0.328	7.64e-4
a_0	2.49e-6	4.67e-6	0.533	3.22e-6	7.89e-7	4.082
a_1	0.894	1.02e-2	87.999	0.874	3.55e-2	24.600
b_1	0.036	9.95e-3	3.610	0.038	3.54e-3	10.603
c_1	0.070	1.12e-2	6.226	0.089	1.64e-2	5.426

Table 4 GJR-GARCH(1,1) model estimation results for return series of S&P 500 and FTSE

	ATX			PX		
	coeff	S.E.	z-score	coeff	S.E.	z-score
α	138.37	12.919	10.711	160.74	12.437	12.924
β	-3.929	0.930	-4.227	-4.265	0.345	-12.351
μ	9.07e-4	8.80e-4	1.031	9.51e-4	1.31e-4	7.256
a_0	4.68e-6	1.18e-6	3.957	4.14e-6	1.16e-6	3.568
a_1	0.891	2.06e-2	43.267	0.868	1.82e-2	47.821
b_1	0.032	9.20e-3	3.478	0.072	1.17e-2	6.130
c_1	0.077	1.71e-2	4.481	0.0602	1.36e-2	4.425

Table 5 GJR-GARCH(1,1) model estimation results for return series of ATX and PX

The estimation results in Tables 2-5 show that leaving the estimation result for μ aside, according to z-score all estimates are statistically significant (different from zero) except the estimate of a_0 for returns of index S&P 500 in model GJR-GARCH(1,1) and the estimate of β for returns of index FTSE 100 in model GARCH(1,1). In model GJR-GARCH(1,1) all estimates of coefficient c_1 are statistically significant indicating that the leverage effect present in all series are statistically significant. Regarding the estimates of parameter μ, they tend to be statistically insignificant. It means that the location parameter of of the NIG distribution of all returns series are not statistically significantly different from zeros. It does not mean that the mean of the returns of these stock market indices are zero. However, it indicates the the distribution of returns of these indices tends to be symmetric.

The next step in this research is to compare the performance of GARCH(1,1) and GJR-GARCH(1,1) models with various distribution assumptions. These assumptions are: normal distribution, Student t-distribution and the NIG distribution. First, the same logarithmic return series are used to estimate parameters of GARCH(1,1) and GJR-GARCH(1,1) models with normal and Student t-distribution specification. For this purpose the two built-in Matlab function garch and gjr are used. These estimates and the estimates of GARCH(1,1) and GJR-GARCH(1,1) models with the NIG distribution obtained in the previous section are used to generate series of conditional variance for the corresponding return series. The conditional variance series are then compared with the corresponding series of squared returns. The results are displayed in Figures 1 and 2 for GARCH(1,1) model and in Figures 3 and 4 for GJR-GARCH(1,1) model.

[3]Z-scores are computed using asymptotic properties of estimates obtained by MLE method. They are test statistics of the null hypothesis suggesting that the estimates are zero.

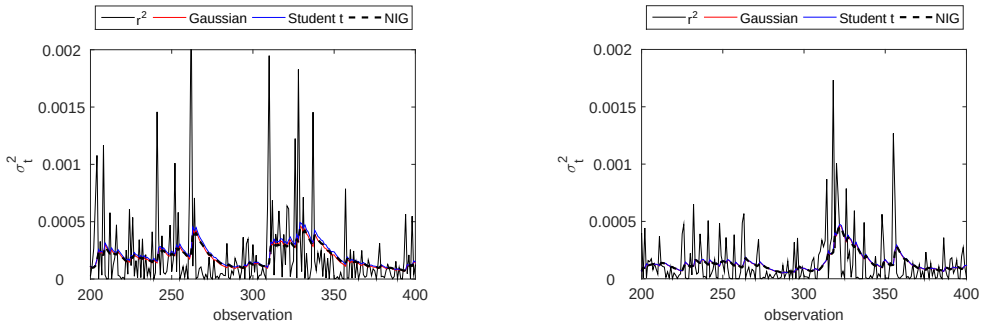


Figure 1 GARCH(1,1) performance comparison for returns of index S&P (left) and FTSE (right)

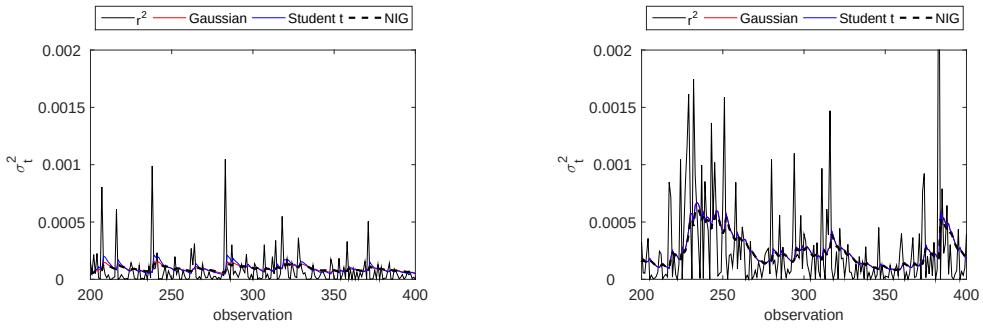


Figure 2 GARCH(1,1) performance comparison for returns of index ATX (left) and PX (right)

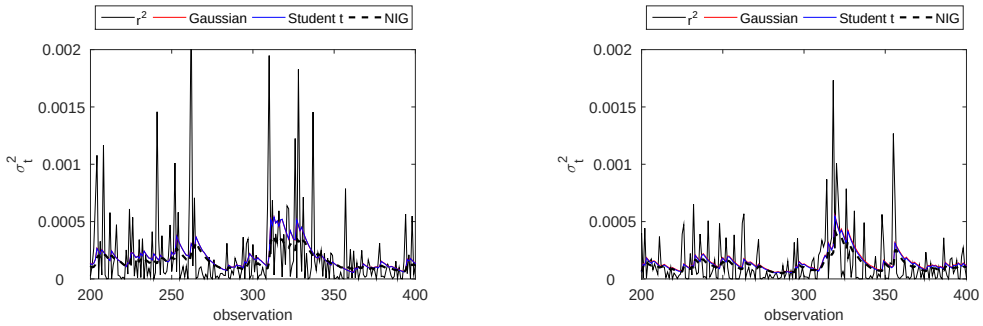


Figure 3 GJR-GARCH(1,1) performance comparison for returns of index S&P (left) and FTSE (right)

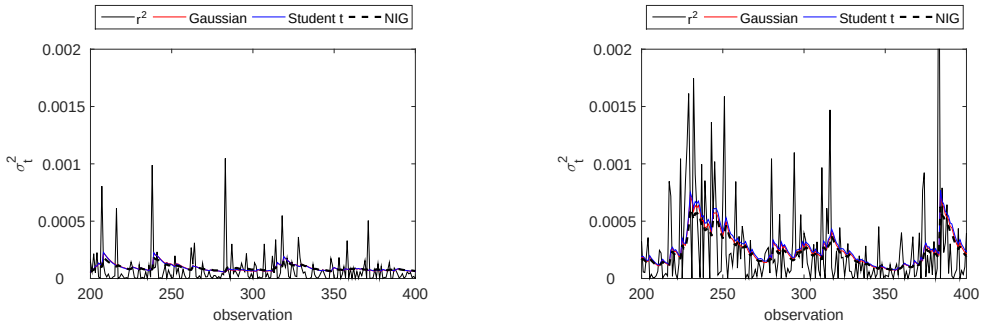


Figure 4 GJR-GARCH(1,1) performance comparison for returns of index ATX (left) and PX (right)

The visual comparison in Figures 1 and 2 for GARCH(1,1) model indicates that there is no substantial difference in conditional variance series generated by a GARCH(1,1) model with normal, Student t and NIG distributions. With t-distribution, the conditional variance seems to be slightly higher than with the other two. With the GJR-GARCH(1,1) model, the variance series generated by this model with normal and Student t distributions are

still close to each other while the series generated with the NIG distribution is visibly lower than the first two. The reason for it may come from the fact that the values of the estimates with normal and t-distribution are higher and their sum is higher than one while the sum of all estimates of this model with NIG distribution still satisfies condition of being less than one. This discrepancy among three series may disappear if this condition is reinforced for the model with normal and t-distribution. Therefore, in our opinion, a GARCH(1,1) model or GJR-GARCH(1,1) model with the NIG distribution for the innovation terms may not be worse than their counterparts with normal distribution or Student t-distribution. On the other hand, they seem not to be better either.

5 Conclusion

In this paper, we have investigated the suitability of the use of normal inverse gaussian distribution for modeling the leptocurticity and volatility clustering properties present in financial asset returns as an alternative distribution to normal distribution and Student t-distribution. To achieve this objective, we have used daily return series of four stock market indices: S&P 500 (USA), FTSE 100 (Great Britain), ATX (Austria), and PX (Czech Republic) and estimated parameters of two most often used models for volatility clustering modeling: GARCH model and GJR-GARCH model. The estimates together with those of models with normal distribution and Student t-distribution have then been used to generate corresponding series of conditional variance for the chosen data. Comparing these series of the corresponding model for each index together, we have found that there is no visible difference among series of conditional variance in the case of the GARCH(1,1) model. In the case of the GJR-GARCH(1,1) model, we have observed a palpable divergence between series with NIG distribution and series with normal and t-distribution. The cause of this divergence may be the inflation of the estimates with normal and t-distribution whose sum is higher than one. In our opinion, in order to competently determine which distribution fits best for a GARCH model, some systematic pre-estimation as well as post-estimation hypothesis testing needs to be done.

Acknowledgements

The financial support from the Czech Science Foundation under Grant P402/12/G097 is gratefully acknowledged.

References

[1] Andersson, J. (2001). On the Normal Inverse Gaussian Stochastic Volatility Model. *Journal of Business and Economic Statistics* 19(1), 44—54.

[2] Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. In *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, The Royal Society 353 (1674), 401–409.

[3] Barndorff-Nielsen, O. E. (1997). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics* 24, 1—13.

[4] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307—327.

[5] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50(4), 987–1007.

[6] Forsberg, L. & Bollerslev, T. (2002). Bridging the Gap Between the Distribution of Realized (ECU) Volatility and ARCH Modeling (of the EURO): The GARCH Normal Inverse Gaussian Model. *Journal of Applied Econometrics* 17, 535—548.

[7] Glosten, L., Jagannathan, R. & Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1179–801.

[8] Guo, Z.-Y. (2017). Heavy-tailed Distributions and Risk Management of Equity Market Tail Events. *Journal of Risk & Control* 4(1), 31-41.

[9] Hansen, P. R., Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1). *Journal of Applied Econometrics* 20, 873–889.

[10] Jensen, M. B. & Lunde, A. (2001). The NIG-S&ARCH model: a fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econometrics Journal* 4, 319—342.

[11] Kilic, R. (2007). Conditional Volatility and Distribution of Exchange Rates: GARCH and FIGARCH Models with NIG Distribution. *Studies in Nonlinear Dynamics & Econometrics* 11(3), 1430-1463.

[12] Stentoft, L. (2008). American Option Pricing Using GARCH Models and the Normal Inverse Gaussian Distribution. Journal of Financial Econometrics, 6(4), 540-582.

Regional analysis: Diversity of neighbourhood relation specification

Simona Mackova[1]

Abstract. Spatial econometrics represents a useful approach, which should not be omitted whilst analysing macroeconomic regional data. Relations are estimated regarding the geographical location of each unit. Spatial econometric models use neighbourhood relations represented by a spatial weight matrix - one of the key elements of the spatial analysis. The spatial weight matrix is non-stochastic, it cannot be estimated within the model and has to be specified at the beginning of the analysis. Since there is no unambiguous procedure to determine the spatial weight matrix and several approaches are available, it is up to the analytic, which approach will be used. Selecting incorrect specification might have severe impact on the result. Therefore aim of this contribution is to empirically test various possibilities of the neighbourhood specification and observes stability of estimated coefficient. Robustness of appropriate model with respect to the neighbourhood specification is evaluated. Analysis is performed on NUTS2-level regions of western European countries. Cross-section macroeconomic regional data describing gross domestic product, unemployment rate and disposal household income are evaluated.

Keywords: spatial econometrics, spatial weight matrix, regional analysis, model robustness.

JEL classification: C21, C52, E66
AMS classification: 91B72

1 Introduction

Spatial econometrics employs models that can be useful for analysis of regional data. This paper observes macroeconomic application of this method. Variables can be influenced by the geographical location of observed unit. If the spatial relations are not involved in the analysis, some important relations can be neglected in the estimation. For spatial regression, so called *geo-coded data* are necessary. This fact could be limiting in choice of suitable datasets.

Neighbourhood relation is defined among regional units by spatial weight matrix. This contribution focuses on neighbourhood specification and weight matrix construction and evaluates impact on estimated coefficients. Matrix $\boldsymbol{W} = [w_{ij}]$ for $i, j = 1, 2, \ldots, N$, where N denotes number of regions, is an important element of spatial econometrics. It should be composed regarding to the observed environment, distribution and size of regions, since it may have impact on results.

The fact, that the matrix $\boldsymbol{W}$ is non-stochastic, is considered as a weakness of spatial modelling in academic literature. $\boldsymbol{W}$ cannot be estimated and has to be defined before other further steps of the analysis. There is no unambiguous solution how to determine a spatial weight matrix. Robustness with respect to the matrix $\boldsymbol{W}$ specification can be verified as suggested in Formánek and Hušek [5].

As an empirical approach of spatial weight matrix construction, method based on Monte-Carlo simulation is suggested in article Stakhovych and Bijmolt [8]. Results can be compared based on information criterion or Log-likelihood function. This process is also criticized for finding only local maximum. Only limited number of matrices is compared and the right specified might not be among them. Incorrectly defined matrix $\boldsymbol{W}$ could cause distortion of estimated coefficients. Otherwise, if the spatial dependence is strong, the results should not be distorted. LeSage and Pace [7] provide theory that strong sensitivity of the model to the specification of spatial weight matrix is just a myth based on the misinterpreted spatial regression estimates or model misspecification.

There is a wide range of spatial models. The full model for cross-section data, which describes all possible spatial relations, is denoted as GNS model (general nesting spatial). Variants of this model can be built by omitting particular spatial interactions. Possible cross-section models are described in detail in Elhorst [3].

This paper is structured as follows: Section two covers neighbourhood specifications and spatial weight matrix derivation, provides theoretic basis and considers possibilities of neighbourhood relations. Section three denotes

[1]University of Economics, Prague, Department of Econometrics, Nam. W. Churchilla 4, 130 67 Prague, Czech Republic, macs03@vse.cz

a chosen spatial model and used formulas applied on real data in section four. Illustrative application and comparison of used neighbourhood specification methods are performed. The last section concludes this paper and lists used references.

2 Spatial weight matrix

2.1 Neighbourhood specification

First, the neighbourhood matrix $\boldsymbol{S}$ is specified. Elements of matrix are zero or one. Number one stands for neighbourhood relation between the two units. Number zero denotes, that units are not neighbours. Diagonal elements are also zeros: the unit is not considered as a neighbour to itself. Matrix $\boldsymbol{S}$ may or may not be symmetric. This depends on the definition of neighbourhood specification. For simplification, only symmetric matrices will be considered in further analysis.

Let us suppose working with real regional geo-coded data: longitude and latitude is known. The most intuitive method of determining neighbours is the method of common border: *contiguity method*. This method can be generalized, and orders of neighbours can be taken into account, e.g: the second order neighbour of an observed unit is a direct neighbour of a first order neighbour of the observed unit. Further, neighbours can be weighted according to their order, e.g. the matrix $\boldsymbol{S}$ contains one half instead of number one for the second order neighbour.

Second suggested approach is *distance-based*: method of maximal regional distance *d*, that is measured by the distance between the so called *centroids*. The centroid can be the capital of the region, administrative unit, important traffic hub or simply the geographic center. Choice from the suggested approaches usually depends on details of available dataset. The distance between two centroids is measured by Euclidean distance. For simplification, curvature of the globe and landscape relief is not considered. Heterogeneity of regions is a weakness for this approach. Regions marked as NUTS were introduced by Eurostat for comparison and analysis of economic variables. Distribution of countries is based on number of inhabitants and the area is not considered. Therefore, area of the regions can be very different, same as the distance between centroids. This causes a large number of neighbours for small regions and only a few neighbours for the large ones. Regions without any neighbours are referred as *island*. To avoid generating islands, at least the maximal distance *d* of the closest neighbour for all regions has to be chosen.

To avoid unbalanced number of neighbours *k-nearest neighbours* approach can be chosen; problem with unequal region size is eliminated. Number *k* must be denoted carefully to not to omit some important connections. This method usually generates asymmetric matrix; therefore the neighbourhood relation is hardly interpretable. A symmetric matrix can be easily derived if at least *k* neighbours are required.

Other mentioned approaches are based on assumption that each region is in neighbourhood relation with every other region weighted by the distance between *i*-th and *j*-th unit d_{ij} or its transformation; influence is assumed as inversely proportional to distance. Non-diagonally elements in the matrix $\boldsymbol{S}$ are replaced by the fraction $\frac{1}{d_{ij}}$ to express relation between *i*-th and *j*-th unit. Formánek and Hušek [5] suggest quadratic or logarithmic transformation to decrease or increase weights calculated for more distant neighbours. Proportions like $\frac{1}{d_{ij}^2}$ or $\frac{1}{log(d_{ij})}$ can be assumed.

2.2 Matrix standardization

For further analysis, matrix $\boldsymbol{S}$ is standardized, and spatial weight matrix $\boldsymbol{W}$ is created. Usually row standardization is employed and it applies $w_{ij} \in [0,1]$ and $\sum_{j=1}^{N} d_{ij} = 1, i = 1, 2, \ldots, N$, where N denotes number of observed regions. The row standardization is used for its interpretability. Elhorst [3] suggests also other possibilities. Column standardization, where columns return sum of one, appears in socio-economic literature.

The literature points out a weakness of this process; the spatial weight matrix $\boldsymbol{W}$ is asymmetric, even though the spatial neighbourhood matrix $\boldsymbol{S}$ was symmetric. It implies that impact of the i-th unit to the j-th unit is different to influence of the j-th unit to the i-th unit, which in many cases might not be true.

More complex alternatives and their implementation in the statistical software R are described in Bivand [2].

3 Spatial model estimation

Verification of spatial autocorrelation should be the next step of the analysis. Moran *I* and Geary's *c* are the most commonly used statistics, which are described in Anselin [1]. Once the relation between location of observed unites is significant, the analysis can continue and spatial regression model can be estimated.

Several model specification and estimation methods are available for cross-section data analysis. In this contribution, the GNS model is adjusted and only spatial interaction of dependent variable is taken into account. Spatial interaction of regressors and random part are neglected.

3.1 Spatial lag model

Let us denote SAR as a spatial lag model. The spatial lag is formulated as a sum $\sum_{j=1}^{N} w_{ij}Y_j$. Since the matrix $\boldsymbol{W}$ is row standardized, the sum represents weighted average of values in neighbouring regions. The SAR model may be written as

$$\boldsymbol{Y} = \delta\boldsymbol{W}\boldsymbol{Y} + \alpha\boldsymbol{\iota}_N + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{1}$$

Under the assumption of existence of an inverse matrix to $(\boldsymbol{I}_N - \delta\boldsymbol{W})$, following form can be derived

$$\begin{aligned} \boldsymbol{Y} - \delta\boldsymbol{W}\boldsymbol{Y} &= \alpha\boldsymbol{\iota}_N + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \boldsymbol{Y} &= (\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}(\alpha\boldsymbol{\iota}_N + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}). \end{aligned} \tag{2}$$

Vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)^T$ stands for dependent variable and $\boldsymbol{X}$ for matrix of regressors. $\boldsymbol{W}$ is a spatial weight matrix and $\boldsymbol{W}\boldsymbol{Y}$ is an endogenous interaction effect of the dependent variable. $\boldsymbol{I_N}$ represents the $N \times N$ identity matrix and $\boldsymbol{\iota}_N$ the $N \times 1$ identity vector, where N in number of observed units. $\boldsymbol{\epsilon}$ stands for the random part. $k \times 1$ vector, where k is number of regressors, $\boldsymbol{\beta}$ and scalars α as an intercept and δ as a spatial parameter are estimated.

Again, various estimation methods are available. The most used procedure is the method of maximum likelihood. Details on this approach are provided in LeSage and Pace [6].

3.2 Direct and indirect effects

Since the spatial weight matrix $\boldsymbol{W}$ is included in the regression model, estimated parameters cannot be interpreted in the same way as in case of classic linear regression. Interpretation of the estimates is based on direct and indirect effects, also called spillovers. The effects are derived as follows:

$$\begin{aligned} (\boldsymbol{I}_N - \delta\boldsymbol{W})\boldsymbol{Y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \alpha\boldsymbol{\iota}_N + \boldsymbol{u} \\ \boldsymbol{Y} &= (\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta}) + \boldsymbol{R}. \end{aligned} \tag{3}$$

$\boldsymbol{R}$ contains the intercept and the random part. Partial derivative of the expected value $E(\boldsymbol{Y})$ by the k-th regressor $\boldsymbol{X_k}$ for all units $1, 2, \ldots, N$ is performed.

$$\begin{pmatrix} \frac{\partial E(\boldsymbol{Y})}{\partial X_{1k}} & \cdots & \frac{\partial E(\boldsymbol{Y})}{\partial X_{Nk}} \end{pmatrix} = \begin{pmatrix} \frac{\partial E(\boldsymbol{Y}_1)}{\partial X_{1k}} & \cdots & \frac{\partial E(\boldsymbol{Y}_1)}{\partial X_{Nk}} \\ \cdots & \cdots & \cdots \\ \frac{\partial E(\boldsymbol{Y}_N)}{\partial X_{1k}} & \cdots & \frac{\partial E(\boldsymbol{Y}_N)}{\partial X_{Nk}} \end{pmatrix} =$$

$$= (\boldsymbol{I}_N - \boldsymbol{\delta}\boldsymbol{W})^{-1}\boldsymbol{\beta_k} \tag{4}$$

The diagonal elements of the matrix (4) represents the direct effect and the non-diagonal elements the indirect effect. Obviously, the effects are different for each observed unit and k matrices of the size $N \times N$ are derived. This fact complicates making any conclusion. LeSage and Pace [6] suggest approaches of measuring for all units. The overall direct effect is calculated as average of diagonal elements. The overall indirect effect is defined as average of sum of row or column of non-diagonals elements. In general, the indirect effect can be interpreted as impact of change of a particular regressor to the dependent variable of all units.

4 Application on western European counties

This section focuses on estimation of a SAR model. Dependence of disposable household income on GDP per capita and on unemployment rate is assumed and evaluated on Western European countries: Belgium, France, Germany, Luxembourg and Netherlands are used for illustrative application.

4.1 Data and model

For this study, NUTS2-level regions are used. All the variables are available from Eurostat database [2]. The observed countries are divided into 83 regions. NUTS2 regions have number of inhabitants between 800 000 and 3 000 000. Datasets for GDP, disposal household income and unemployment rate are directly downloaded using statistical software R and its packages. Values from the year 2014 are analysed. Data are available in tables under following codes: nama_10r_2gdp, nama_10r_2hhinc, tgs00010.

Let us denote gross domestic product per capita as ***GDP***, unemployment rate as ***UNE*** and disposal household income as ***INC***. The appropriate SAR model is defined as

$$INC = \alpha + \delta \boldsymbol{W} \cdot INC + \beta_1 GDP + \beta_2 UNE + \boldsymbol{\epsilon}. \tag{5}$$

4.2 Neighbourhood specification

Five variants of the spatial weight matrix $\boldsymbol{W}$ are introduced in this section. The null hypothesis of absence of spatial autocorrelation was rejected in all cases; spatial relations should be considered. For estimations and spatial weight matrix construction, software R in employed.

First, the k-nearest neighbours approach is used. Algorithm for choice of the right k is inspired by Formánek and Hušek [5]. Different k between 1 to 80 were used. The one which minimized the Akaike information criterion (AIC) and maximized the Log-likelihood function was chosen for further comparison. Results of this partial analysis is summarized in figure 1. k equal to six minimizes the AIC statistics. $\boldsymbol{W_1}$ denotes the spatial weight matrix assuming six nearest neighbours.

Figure 2 shows results of analogical approach applied on maximal centroid distance. Threshold of 180 km guarantees existence of at least one neighbour for each unit and non-existence of islands. The maximal distance is increased by 10 km until 1380 km. The lowest maximal distance of 180 km minimized AIC. The matrix $\boldsymbol{W_2}$ is derived based on this fact.

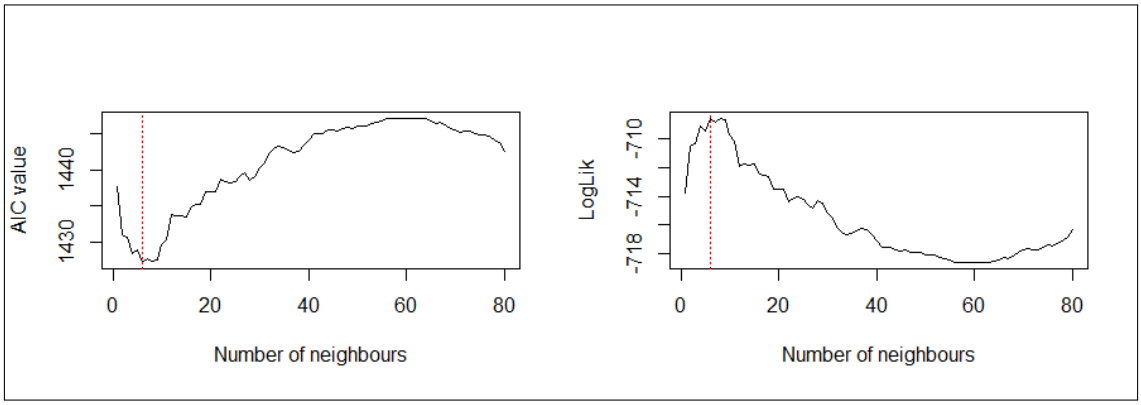


Figure 1 AIC and Log-likelihood development dependent on maximal neighbours distance in km

The third suggested approach is based on common borders of regions. This process generates the most balanced number of neighbours in the whole dataset. Let us denote the related matrix as $\boldsymbol{W_3}$.

The last two possibilities weight the neighbourhood relation by the distance and by the logarithm of the centroid distance. Normalized matrix $\boldsymbol{W_4}$ has zeros on the diagonal and $w_{ij}, i \neq j$ is defined as

$$w_{ij} = \frac{d_{ij}^{-1}}{\sum_{k=1}^{N} d_{ik}^{-1}}. \tag{6}$$

Off-diagonal elements of the weight matrix $\boldsymbol{W_5}$ are

$$w_{ij} = \frac{[log(d_{ij})]^{-1}}{\sum_{k=1}^{N} [log(d_{ik})]^{-1}}. \tag{7}$$

[2] http://ec.europa.eu/eurostat/data/database

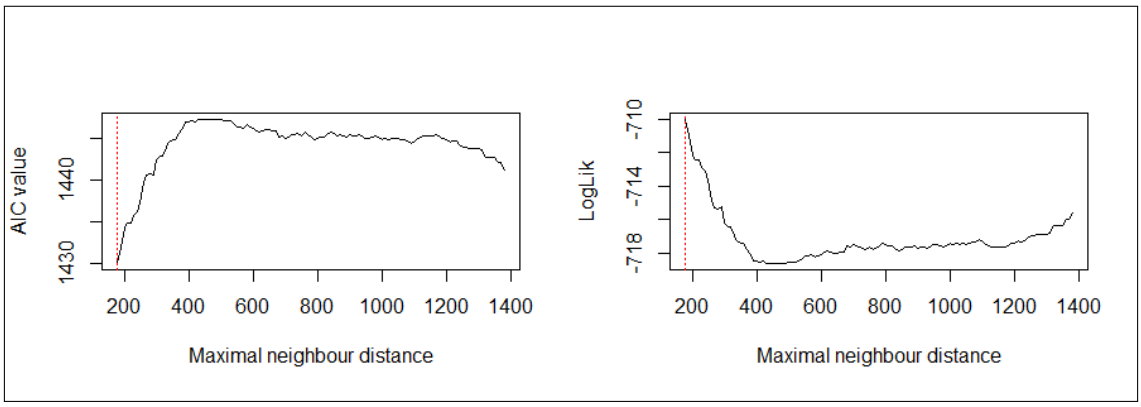


Figure 2 AIC and Log-Likelihood development dependent on number of neighbours

For all these introduced spatial weight matrices, SAR model as in equation 5 is estimated by the method of maximum likelihood.

4.3 Comparison of results

Model is supposed to be robust with respect to the spatial weight matrix changes. Change of the neighbourhood specification method should not have significant impact on estimated coefficients and effects. See the table 1, where each row belongs to one of the weight matrices and quantifies statistic Moran I, parameter δ estimate, direct, indirect and total effect of both regressors: *GDP* and *UNE*. Statistical significance level is marked.

	Moran I	δ	**Direct effect**	**Indirect effect**	**Total effect**	**AIC**	
W_1	0.340**	0.469**	0.12853**	0.10365**	0.23218**	1 427.1	*GDP*
			-296.920**	-239.437**	-536.358**		*UNE*
W_2	0.327**	0.442**	0.12553**	0.09098**	0.21657**	1 430.0	*GDP*
			-312.248**	-226.182**	-538.431**		*UNE*
W_3	0.352**	0.174*	0.11447**	0.02317	0.13764**	1 444.1	*GDP*
			-364.346**	-73.747*	-438.093**		*UNE*
W_4	0.002*	0.721**	0.12381**	0.30992	0.43373	1 442.1	*GDP*
			-424.769**	-1 063.276	-1 488.045		*UNE*
W_5	-0.010**	0.003	0.11916**	0.00040	0.11955	1 447.2	*GDP*
			-402.649**	-1.330	-403.979		*UNE*

** Significant at 5%, * Significant at 10%

Table 1 Estimated coefficients end effects of the SAR models with various matrices W

The first two methods seem replaceable. Estimates of the spatial parameter δ do not differ much as well as all the estimated effects. The last two methods based on inverse provided very similar direct effects. However, their indirect effects are not only far different but also statistically insignificant. One can also look at correlations between the spatial weight matrices. Matrices with high correlation are supposed to deliver similar results as claimed in LeSage and Pace [7]. The table 2 validates this premise. The highest correlation coefficient is between matrices W_4 and W_5. Also coefficients higher than 0.5 are between W_1 and W_2 and between W_1 and W_3.

	W_1	W_2	W_3	W_4	W_5
W_1	-	0.562	0.501	0.004	0.020
W_2	0.562	-	0.336	0.014	0.028
W_3	0.501	0.336	-	0.002	0.017
W_4	0.004	0.014	0.002	-	0.792
W_5	0.020	0.028	0.017	0.792	-

Table 2 Correlation coefficients between the various matrices W

5 Conclusion

This paper provides basic theory for spatial econometrics focused on the neighbourhood specification and the weight matrix calculation. Several approaches were suggested and applied on data for Western European countries. According to the illustrative example, statements from the article LeSage and Pace [7] can be confirmed. First, even from the equation 1 can be clearly seen, that the spatial weight matrix $\boldsymbol{W}$ plays a role in spatial econometrics and its variants do not perform equally. Very different matrices can bring materially different results and sensitivity of the estimates was observed. Second, correlation coefficients validated that some variants of matrices $\boldsymbol{W}$ share common elements and their results can be highly similar. In this contribution, the spatial structure showed as an important parameter of spatial regression. On the other hand, some variants of matrix $\boldsymbol{W}$ can bring similar results and there is no need to put too much effort in searching for the best possibility.

Acknowledgements

This research project was supported by the grant No. IGA F4/60/2018, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Anselin, L. (2003). An Introduction to Spatial Regression Analysis in R. Illinois: *Urbana-Champaign*.

[2] Bivand, R. (2016). Creating Neighbours. URL `https://cran.r-project.org/web/packages/spdep/vignettes/nb.pdf`.

[3] Elhorst, J. P. (2014). *Spatial econometrics: from cross-section data to spatial panel*. Heidelberg: Springer.

[4] Formánek, T. & Hušek, R. (2015). The Czech Republic and its neighbors: Analysis of spatial macroeconomic dynamics. In 33th International Conference Mathematical Methods in Economics 2015, *MME* (pp. 190-195).

[5] Formánek, T. & Hušek, R. (2016). On the stability of spatial econometric models: Application to the Czech Republic and its neighbors. In 34th International Conference Mathematical Methods in Economics 2016, *MME* (pp. 213-218).

[6] LeSage, J. & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. U.S.A.: CRC Press.

[7] LeSage, J. P. & Pace, J. P. (2014). The biggest myth in spatial econometrics. *Econometrics*, *2*(4), (pp. 217-249).

[8] Stakhovych, S. & Bijmolt, T. H. A. (2009). Specification of spatial models: A simulation study on weights matrices. *Paper in Regional Science*, (88.2), (pp. 389-408).

What is the most appropriate characteristic of wages?

Lubos Marek[1]

Abstract. There are two questions that interest almost everyone. What is my wage? How am I doing compared to others? The arithmetic average is used as the usual characterization of the wage level. Everyone knows or suspects that this average is far from ideal. It is generally known that about 1/3 of wages are above the average, that is, the remaining 2/3 of wages are below it. So the question arises why other (better?) indicators of wages are not being published. We mean the median or other types of average, for example. We will try to show in this paper the other characteristics which could be used as useful indicators of wage levels. We apply these characteristics to real data and compare the results achieved. We work with a large data set (over two million observations). Our data covers years 1995-2016. All of our calculations will be carried out on this data. We also assess the advantages and disadvantages of different approaches. We have disaggregated the wage data by various factors (gender, age, region, education, and others).
This enables us to perform interesting comparisons not only for the entire country, but also by such factors. The entire analysis is carried out in MS Excel.

Keywords: wage, wage characteristics, arithmetic average, median.

JEL Classification: C44
AMS Classification: 90C15

1 Wages in the Czech Republic

The arithmetic mean is the most frequently used characteristic of wages. The average wage value is one of the most-closely watched economic variables by both the general public and economists. Questions "How much do I earn?" and "What is my income in comparison with the others?" are interesting for each and every employee. The arithmetic mean is certainly not an ideal characteristic of wages. We all know that it is sensitive with respect to influence by outliers. Hence we can ask whether there exists a better characteristic or if the arithmetic means should be complemented by other suitably selected characteristics. In this paper we will calculate some other wage characteristics and compare them with the traditional average.

Our analysis will be carried out on a large data set– about two million observations per year, for years 1995-2017, always the average wages in the second quarter of the given year. The data is given in the form of interval frequencies with an interval width of 500 CZK. In other words, it is a rather detailed data set, which will enable us to calculate all the necessary characteristics. The source of our data is the Trexima Company– [7].

1.1 Empirical Distribution of Wages

Let us first have a look at the time evolution of the wage distribution in the Czech Republic. Figure 1 shows the empirical distribution (i.e., polygon) of wages. The chart implies that all of the basic statistical characteristics of wages have been undergoing changes; this concerns position, variability, skewness, and kurtosis. In other words, the average wage values are growing, individual wage values are more and more distinguished from each other, and the positive skewness of the wage distribution is growing, while its kurtosis is decreasing. We are most interested in the shifting position, represented in by the arithmetic mean (i.e., average value) of wages in Figure 1.

Figure 2 shows the time evolution of the average wage value, which is a more or less linear growth. If we capture the trend with a line, the estimated trend (with the respective equation written directly in the Figure) with the determination index equal to 0.9873, which means a very tight linear dependency. An interesting point is that the average wage values were growing even during the economic crisis, which is rather unexpected. On the other hand, the wages are expressed in current prices, not cleaned of the inflation rate effects. Had we incorporated the inflation rate into our considerations, the situation would look different: the growth at the time of crisis would be

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, nam. W. Churchilla 4, 130 67 Praha 3, marek@vse.cz.

replaced with stagnation or a slight decrease. This aspect is addressed in Figure 3, in which the average wage values are depicted with respect to the inflation rate – related to both 1995 and 2017 price levels.

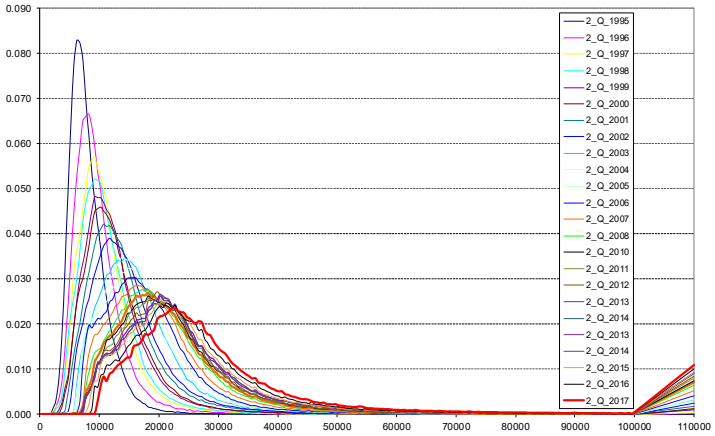


Figure 1 Polygon of wages

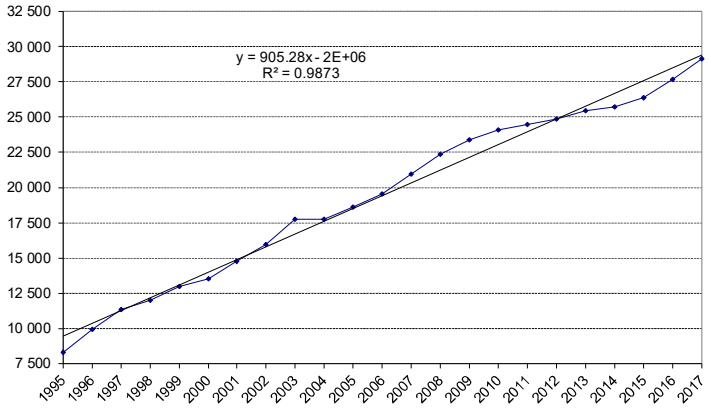


Figure 2 Average wage values

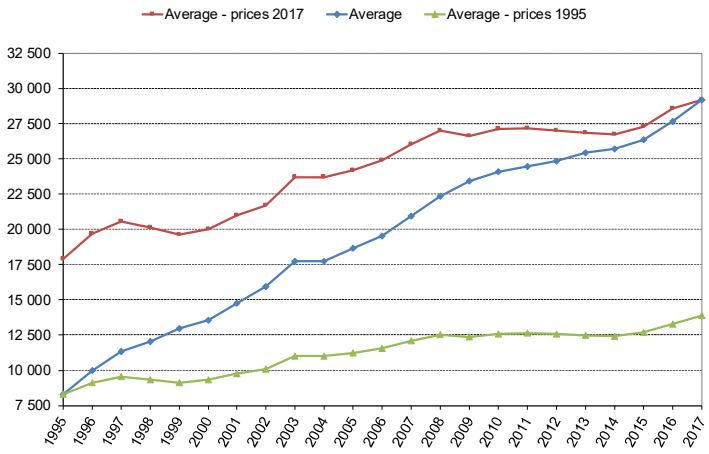


Figure 3 Average wage values with inflation effects

Let us return to Figure 1. The tail of this distribution corresponds to high wages, There are not too many of those, but they significantly affect the average values – cf. [1], [4], and [5] for more details. These high wages play the roles of outliers here, which increase the average value. That is one of the reasons why 2/3 of the population is below the average wage level.

1.2 Wage Characteristics

Let us now propose some other characteristics of wages. We first introduce them and then we calculate their values on our data set, compare them with the average values, and evaluate our results. We will make use of certain additional types of average, and also quantiles – cf. Mala [3]. First of all, it is the so-called *winsorized average*

$$\overline{x}_{w.k} = \frac{(k+1)y_{k+1} + \sum_{i=k+2}^{n-k-1} y_i + (k+1)y_{n-k}}{n}, \tag{1}$$

where k is a suitable chosen constant to eliminate the outliers. The k value could be specified as percentage, in which case Formula (1) would have to be modified accordingly. Figure 1 indicates that the outliers are mainly present on the right-hand side of the distribution (at the wages above 10,000 CZK), and no such problem occurs on the left-hand side. That is why we suggest that a modified winsorized average value should be used (a one-sided version of Formula (1)), which responds to the distinct skewness of the empirical distribution and the heavy tail of the distribution

$$\overline{x}_{w+.k} = \frac{\sum_{i=1}^{n-k-1} y_i + (k+1)y_{n-k}}{n}. \tag{2}$$

Another option is the *truncated average*

$$\overline{x}_{t.k} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} y_i . \tag{3}$$

It is the arithmetic mean from which a selected number of smallest and largest values are left out. If k observations are left out, the arithmetic mean is calculated from the remaining $n-2k$ values. Similar to the previous instance, we suggest that Formula (3) should be modified to

$$\overline{x}_{t.k} = \frac{1}{n-k} \sum_{i=1}^{n-k} y_i . \tag{4}$$

in which only k largest values are left out, that is, the sample is only truncated from the right. This Formula could be one-sided from the other side (leaving out only k smallest values) if the skewness were negative; however, this is not our present case.

Other characteristics, based on quantiles, can also be used. In this paper, we will use the *median* and the so-called *trimean*

$$\overline{x}_{trimean} = \frac{x_{0,25} + 2x_{0,5} + x_{0,75}}{4}. \tag{5}$$

Formula (5) is a weighted arithmetic mean of the upper and lower quartiles, and the median, with the median's weight doubled with respect to the quartiles' weights, i.e., the median has a weight of 0.5, and the quartiles each a weight of 0.25 in Formula (5).

Theoretical considerations admit other characteristics of the position based on the so-called L-moments [2], or M-estimators [6]; we will not consider them in this paper, mainly due to the space limitations. The respective details can be found in the cited references.

1.3 Results

First, we look at the basic characteristics (the arithmetic mean, median, and trimean) as calculated in the entire period under consideration and shown in Figure 4.

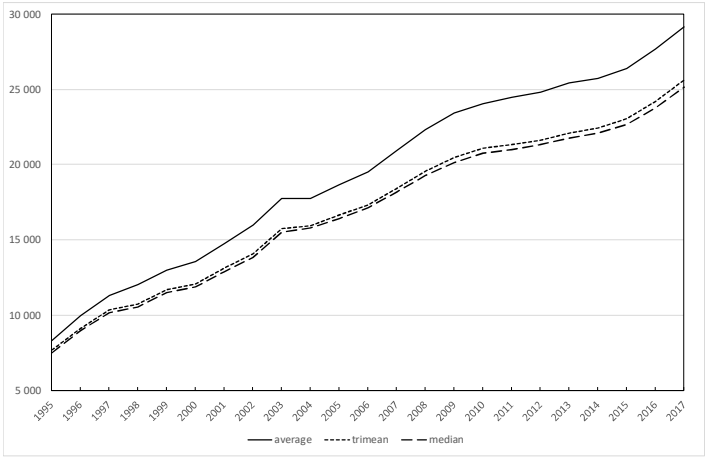


Figure 4 Average, median, and trimean of wages

We can observe that, as expected, the average of wages is the largest. Both the median and the trimean are significantly smaller, the latter somewhat larger than the former. The difference between each pair of characteristics is growing with time.

Let us consider the winsorized (1) and truncated (3) averages for different values of k in 2017. For comparison, we also refer to the numerical values used in Figure 4.

Level	k	$\bar{x}_{w.k}$	$\bar{x}_{t.k}$
100,000	23,795	28,553	27,956
90,000	31,609	28,466	27,813
80,000	43,474	28,275	27,556
70,000	62,485	28,066	27,289
60,000	94,091	27,801	27,017
50,000	154,602	27,304	26,551
46,000	200,000	27,097	26,390

Table 1 Winsorized and truncated averages in 2017

Average	Median	Trimean
29,166	25,135	25,607

Table 2 Average, median, and trimean in 2017

The k value for both average variants in Table 1 is derived from this Table's first column, in which the values are taken so that all observations above a certain limit are eliminated. E.g., for the first value of such level, 100,000, there were 23,795 observations above this level.

With a growing k value, the difference between both variants of the average is growing, as well as their distance from the average value over the Czech Republic as a whole in Table 2. The truncated average's values are always smaller than those of the winsorized one. This fact can be expected from the theoretical values of both these variants – cf. Formulae (1) and (3). At the same time, the k value is not very large; the sample size in 2017 amounts to 2,185,573 observations. On the other hand, the winsorized average value is more robust – it very slowly responds to the growing value of k. The values of both these average variants are much larger than the median value, as well as the trimean value, for all values of k.

Let us re-calculate the winsorized (Formula (2)) and truncated (Formula (4)) average values, this time in their one-sided modifications. The results are summed up in Table 3. With a growing k value, both variants' average values are decreased in compliance with the theoretical considerations. The truncated average values are again

smaller than those of the winsorized average values, and both variants are smaller than the non-modified ones. The latter fact is a consequence of a significant skewness of the wage distribution towards the left-hand side – cf. Figure 1. The modified winsorized average value is still higher than the median and the trimean values; the modified truncated average value is smaller than the median value for k = 200,000.

Level	k	$\bar{x}_{w.k}$	$\bar{x}_{t.k}$
100,000	23,795	28,530	27,743
90,000	31,609	28,404	27,500
80,000	43,474	28,235	27,184
70,000	62,485	27,996	26,760
60,000	94,091	27,644	26,188
50,000	154,602	27,092	25,348
46,000	200,000	26,836	24,931

Table 3 Modified winsorized and truncated averages in 2017

In Figure 5 we compare the values of all the considered characteristics for k = 100,000. Wages above 100,000 CZK play the key role in influencing the average wages. If we disregarded them (in the modified truncated average), the average wage value in the Czech Republic would go down from 29,166 CZK to 27,743 CZK in the second quarter of 2017.

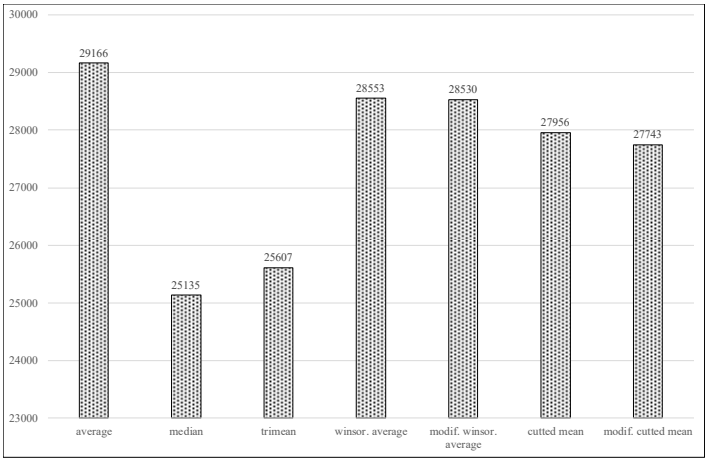


Figure 5 Basic characteristics of wages in 2017

Conclusions

On the basis of our results we can confirm that the arithmetic mean is not ideal for characterizing the wage values because it is distinctly affected by high wages above 100,000 CZK, whose quantity is not large (1.09%) but whose effect on the average value is rather significant (referring to the 2017 data). If we exclude from the calculation the wages above 100,000 CZK by calculating the modified truncated average value, the average wage value goes down by 1,423 CZK, i.e., 4.8%. For all other types of average and quantile values, the results are smaller than the average, sometimes quite substantially. None of them is, however, closer to the median than to the arithmetic mean. That is, none of the four modified average values considered in this paper can be taken for a sole best characteristic of the wage values. As the most suitable approach we should publish the average wage values together with other characteristics – in particular, the median as well as other characteristics, such as other quantiles, or modified winsorized or truncated averages.

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] Doucek, P. & Marek, L. (2016): Vývoj mezd a příjmové nerovnosti u ICT odborníků v České republice [Evolution of wages and inequalities in income among the ICT experts in the Czech Republic]. *Politická ekonomie*. 2016, roč. 64, č. 8, s. 922–938. ISSN 0032-3233.

[2] Malá, I. (2017) L-moments of a finite mixture of probability distributions. In: International Days of Statistics and Economics (MSED) [online]. Praha, 14.09.2017 – 16.09.2017. Slaný : Melandrium, Libuše Macáková, 2017, s. 929–938. ISBN 978-80-87990-12-4. Bilingsley, P. (1968). *Convergence of probability measures*. New York: John Wiley & Sons, Inc.

[3] Malá, I. (2017) Vybrané kapitoly z teorie pravděpodobnosti a matematické statistiky [Selected chapters from theory of probability and mathematical statistics]. 1. vyd. Praha : Nakladatelství Oeconomica, 2017. 230 s. ISBN 978-80-245-2184-8.

[4] Marek, L. (2016): Comparison of wages in the Czech Regions. *In: Mathematical Methods in Economics* [CD-ROM]. Liberec, 06.09.2016 – 09.09.2016. Liberec: TU Liberec, 2016, s. 534–539. ISBN 978-80-7494-296-9.

[5] Pacáková, V. & Linda B. & Sipková, L. (2012) Rozdelenie a faktory najvyšších miezd zamestnancov v Slovenskej republike [Distribution and factors of employees' highest wages in the Slovak Repoublic]. *Ekonomický časopis*, Ekonomický ústav SAV a Prognostický ústav CSPV SAV, p. 918, no 9, vol 60/2012, ISSN 0013-3035.

[6] Terek, M. (2016): Odľahlé dáta a charakteristiky polohy v analýzach miezd a príjmov [Outliers and Measures of Location in Analyses of Wages and Incomes], *Revue sociálno-ekonomického rozvoja*, Ekonomická univerzita v Bratislave, ročník 2, 1/2016, pp. 114-126, ISSN 2453-6148.

[7] Trexima, (2016): http://www.trexima.Cz

Bookmakers' Efficiency in English Football Leagues

Patrice Marek[1]

Abstract. In 2015, the online gambling market had a volume of almost 38 billion U.S. dollars, and the largest portion – 48 percent – was held by sports betting. It is expected that online gambling market will be worth of 60 billion U.S. dollars by the year 2020.
Large number of sports bets are placed on football, especially English football leagues. The high competition in this area forces bookmakers to lower their margins, and this leads to a requirement of better – more precise – models of probabilities of possible match results. This paper analyses margins of selected bookmakers over the last years and compare efficiency of their models based on implied probabilities. Results and odds of five top level English leagues since 2005/2006 season are chosen for the analysis and demonstration of basic characteristics development.

Keywords: Fixed odds betting, football, bookmaker, efficiency, margin, analysis, England.

JEL classification: C58, Z29
AMS classification: 94A17, 62P05

1 Introduction

Gambling market is a largely expanding area, especially in its online form. According to Statista [10], the online gambling market had a volume of 20.51 billion U.S. dollars in 2009, and the volume was almost doubled in 2015 to 37.91 billion U.S. dollars. The largest portion – 48 percent – of the volume in 2015 was held by sports betting. The expected volume of the online gambling market by the year 2020 is 59.79 billion U.S. dollars. The gambling market is, therefore, interesting for bookmakers, and the competition is increasing.

This paper studies fixed odds betting, where bookmakers offer odds on possible match results – home team win, draw, and away team win. Bookmakers has their own models to estimate probabilities of match results. These probabilities and chosen value of margin are used to construct odds that are sub-fair, i.e. if a bettor places stakes exactly according to the estimated probabilities on every possible outcome, he will always lose an amount that is equal to the margin of a bookmaker. This can be illustrated on coin flipping. The probabilities for a fair coin are known, and we do not need to estimate them (0.5 for tail and 0.5 for head). The fair odds with no margin would be 2.0 for each outcome (meaning that if we place 1 unit on head, and this will be the result of the coin flipping, we will win 2 units, i.e. net profit of 1 unit). Nevertheless, in the real world, bookmakers offer sub-fair odds; therefore, a bookmaker who calculates with 10% margin would use odds of 1.8 for each outcome. Now, if a bettor places 0.5 unit on tail and 0.5 on head (exactly according to the real probabilities), then the bookmaker accepts 1 unit, and for each possible outcome the bookmaker will pay only 0.9 unit ($0.5 \cdot 1.8$). This secures him 0.1 unit of profit, i.e. 10% margin, or 11.1% markup.

One important effect of increasing competition is decreasing margin of bookmakers. Nevertheless, to reduce margin in fixed-odds betting can be dangerous as bookmakers do not know the real probabilities and use only estimates. This indicates necessity of good models for modelling probabilities of possible match results; threfore, interest of research in this area is increasing. Models used for estimation of outcomes can be tracked back into 1980s. Maher [7] showed that it is possible to use the bivariate Poisson distribution to estimate football match results. His models were improved in many following papers, e.g. Dixon and Coles [3] introduced model that allowed negative dependencies and possibility to discount information from older matches. Both these papers were concerned by football. Karlis and Ntzoufras [5] studied not only football but also water polo, ice hockey was studied in Marek, Šedivá, & Ťoupal [9], Buttrey [1], and Šedivá [11], and tennis was studied by Kovalchik [6]. As can be seen, theory in this area is developing, and it is – with the increasing computing power – possible to obtain more precise estimates of probabilities and to reduce margin to attract new customers.

Che, Feddersen, & Humphreys [2] studied over-round (or, in accounting terms, markup) of two bookmakers, William Hill and Ladbrokes – both are later used in this paper –, from the 2004/05 season to the 2011/12 season in the English Premier League. These two bookmakers operated both betting shops and online bookmaking operation. They found out that, in a response to increased competition, the over-round – and thus also the margin – substantially dropped for the 2008/09 season. The over-round of William Hill and Ladbrokes are also compared

[1]University of West Bohemia, European Centre of Excellence NTIS – New Technologies for Information Society, Faculty of Applied Sciences, Univerzitní 8, Plzeň, Czech Republic, patrke@kma.zcu.cz

with four pure online bookmakers: Bet365, Bwin, Interwetter, and Sportingbet (first three bookmakers are also used in this paper). They found that over-round of these online bookmakers (mixed together as one) declined more slowly but from lower values. Malarić [8] used results of 10 European leagues during seasons 1999/2000, 2000/01, and 2001/02 and obtained average margin of 13% for average odds computed from 15 bookmakers. He showed that the fixed odds market at that time was inefficient.

The main goal of this paper is to demonstrate changes in margins of studied bookmakers and to – indirectly with usage of implied probabilities based on odds – evaluate their models.

2 Data

Results and odds of five top level English football leagues are used for the analysis, i.e. Premier League (level 1, EN_1), English Football League Championship (level 2, EN_2), English Football League One (level 3, EN_3), English Football League Two (level 4, EN_4), and National League (level 5, EN_5). Data obtained from Football-Data [4] comprise results and odds between the 2005/06 season and 2017/18 season. The 2017/18 season in not finished yet, and only results up to March 27, 2018 are used. However, more than 3/4 of the season are played; therefore, even these data are used so that the paper comprises the most actual results. Totally, 17 229 matches are used for the analysis.

Odds of six bookmakers are analysed – Bet365 ($B365$), Bwin (BW), Interwetten (IW), Ladbrokes (LB), BetVictor (VC), and William Hill (WH). Moreover, two fictitious bookmakers are created from odds of these six companies – MAX that uses maximal odds that are found for each outcome, and AVG that uses average odds of listed bookmakers. Odds are available for the most of the matches; however, for some of them, the database Football-Data [4] does not contain odds. Usually, this is for several matches in higher level leagues and for small percent of matches in lower level leagues. The only exception is the 2007/08 season where the database does not contain odds of WH for 1/3 of matches (other bookmakers are not affected and usually all matches are listed with odds). Results in the following part are standardized for one match; therefore, several missing matches will not significantly affect the final results. The only exception is the 2007/08 season for WH, as mentioned before.

3 Methods and Results

First, we need to compute margin and probabilities that are used by a bookmaker for every single match. To do this, we need to specify model that is used by the bookmaker to compute odds. The simple and common model use assumption that the same margin (ζ) is used for each possible outcome of a match (see Malarić et al. [8]), i.e. that the bookmaker computes odds of the home team win (o_H), draw (o_D), and away team win (o_A) according to

$$o_r = \frac{1-\zeta}{p_r}, r = H, D, A, \tag{1}$$

where $p_r, r = H, D, A$ are probabilities estimated by the bookmaker (see Štrumbelj [12] for more models how to calculate probabilities from odds). Equation (1) can be used to compute margin used for a match according to

$$\zeta = 1 - \frac{1}{\sum_r o_r^{-1}}, r = H, D, A. \tag{2}$$

Equations (1) and (2) can be easily used to compute so-called implied probabilities that were used by the bookmaker in each match, i.e. for each match with odds in used database this offers information about margin and probabilities used by the bookmaker.

3.1 Average Margins

Average margins of each bookmaker in each season and league were computed, and results are presented in Figure 1. We can see that the average margin decreased in time, e.g., average margins used by bookmakers in the Premier League 2005/06 season were between 7.33% ($B365$) and 11.10% (WH), and margins in the current Premier League season are between 2.71% (VC) and 5.79% (LB). Margins are usually higher in the lower level leagues, and the highest margins in the 2017/18 season are in the National League (between 5.64% (VC) and 11.54% (IW)). High margins in lower level leagues indicates possible uncertainty of bookmakers' models. Qualities of their models are investigated in the following parts where three measures – inspired by Kovalchik [6] – are used. Results of MAX, a fictitious bookmaker with the highest odds, show that his average margin in the current season is 1.05%; there is also around 20% of matches with negative margin (62 out of 304 played matches; the lowest margin recorded was -4.12%). Two bookmakers, $B365$ and VC, use margin that is usually lower than the average margin (represented by AVG).

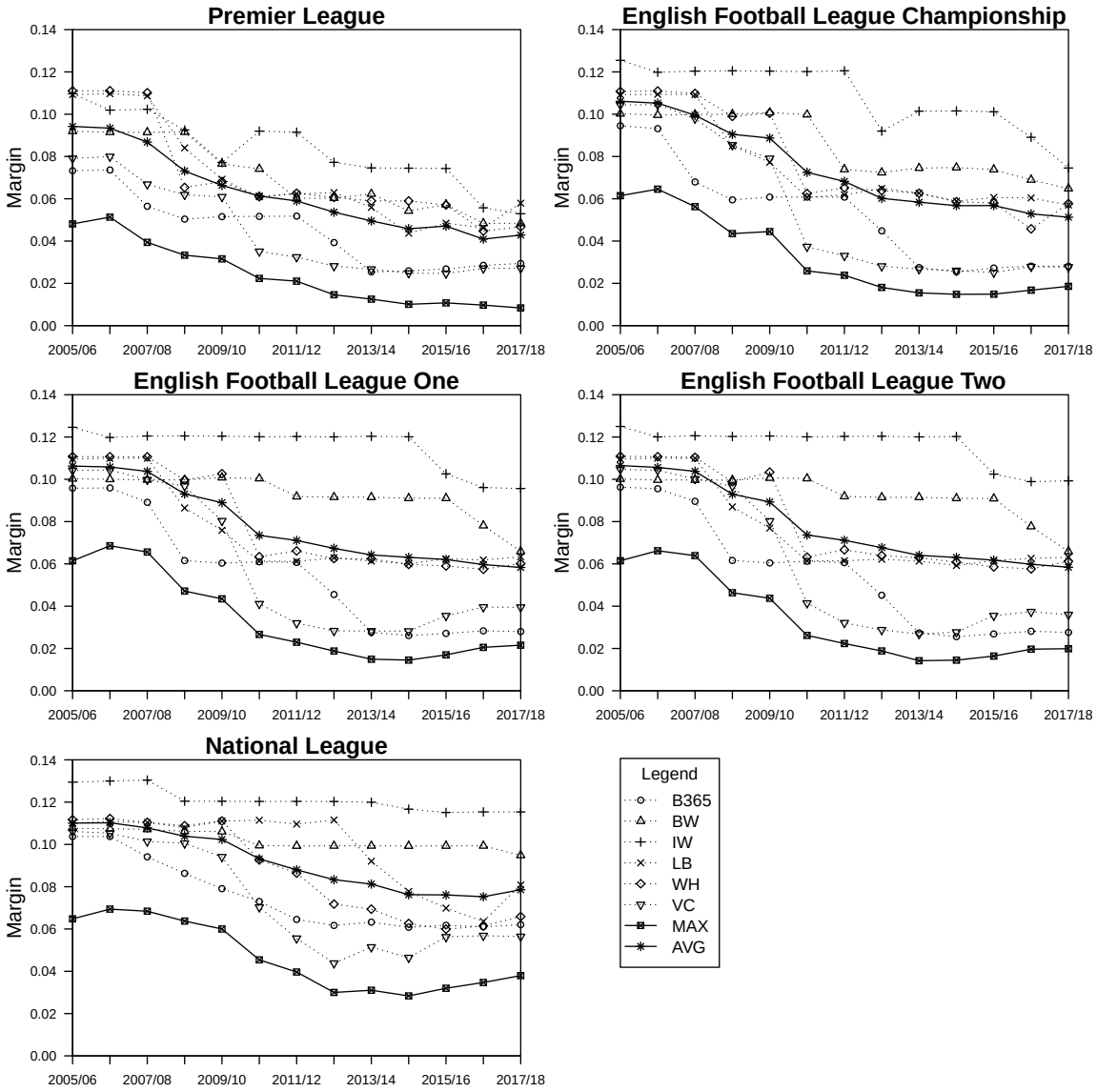


Figure 1 Margins of bookmakers

3.2 Calibration

The calibration is a measure of the model's accuracy which focuses on the number of expected wins. The model is well calibrated when for all matches with the predicted probability p for given outcome the observed proportion is close to p. To decide whether the model is well calibrated or not, the calibration ratio (C) can be used.

$$C = \frac{\sum_{m=1}^{M} \max(p_m^H, p_m^D, p_m^A)}{\sum_{m=1}^{M} \delta_m}, \tag{3}$$

where M is total number of matches, $p_m^r, r = H, D, A$ are probabilities used by the bookmaker (obtained from his odds by Equations (1) and (2)), and $\delta_m = 1$ in the mth match if this match ends with the result that has the highest probability according to the bookmaker, and $\delta_m = 0$ otherwise. In some cases – usually in less than 15 matches of a season –, the number of possible outcomes of a match has to be changed. The first case is when two outcomes have the same probability, and this probability is also equal to the maximum of all probabilities. In this case, these two outcomes are combined together as one with summed probability. The second and very rare case is when all three probabilities in a match are equal, then all three are combined together as one outcome that will happen with probability 1. The most matches in a season with some correction were recorded for WH in EN_4 in the 2014/15 season – 57 out of 552 matches. In the Premier League, the most matches with some correction were recorded in the 2005/06 season for IW – 15 out of 380 matches.

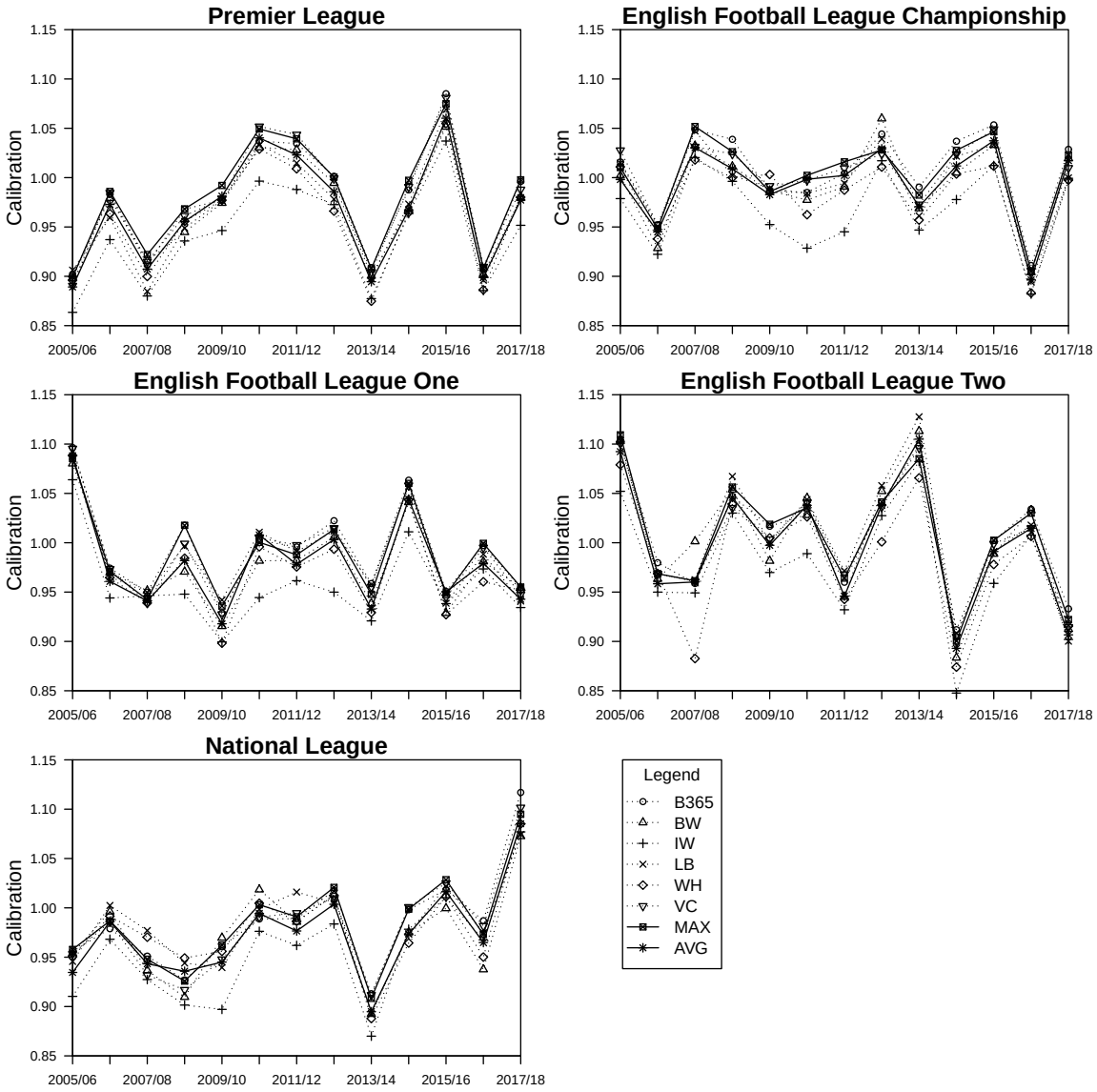


Figure 2 Calibration of bookmakers

Results of calibration are presented in Figure 2. The usual situation is that calibration is around 1 with some exceptions. A situation where the calibration is less than one indicates that probabilities of the most probable outcome are underestimated and vice versa. Figure 2 also indicates that bookmakers use similar models – or at least similar odds – as differences among bookmakers are usually not as high as differences between two seasons. Some minor differences can be identified for IW that has usually the lowest value of calibration, i.e. in the comparison with other bookmaker IW calculates with lower probabilities for the most probable outcome. Next, it is not possible to identify improvements in models as values are still in the same intervals.

3.3 LogLoss Function

The next measure, usually suitable for betting, is $LogLoss$ function. The function for sports with three possible outcomes of a match is defined as

$$LogLoss = -\frac{1}{M}\sum_{m=1}^{M}(\kappa_m^H \ln p_m^H + \kappa_m^D \ln p_m^D + \kappa_m^A \ln p_m^A), \tag{4}$$

where M is total number of matches, $p_m^r, r = H, D, A$ are probabilities used by the bookmaker, and κ_m is an indicator function, e.g., $\kappa_m^H = 1$ if match m ends with a home win, and $\kappa_m^H = 0$ otherwise. This function gives high penalty to situations where outcomes with small predicted probability occur; the lower value of $LogLoss$

function, the better model. $LogLoss$ function is expressed for one match; therefore, leagues with different number of matches in a season can be directly compared.

The highest recorded difference is in the 2007/08 season between WH (1.042) and LB (1.061) in EN_4. Nevertheless, as mentioned before, database contained only 2/3 of odds for WH for this season. If we exclude results of WH in the 2007/08 season, then the highest difference was recorded in the 2008/09 season between IW (1.019) and LB (1.030) in EN_3. These differences are low in the comparison to differences between two consecutive seasons (in some cases, the difference is more than 0.1). Therefore, Figure 3 presents results only of $B365$ (chosen because it is one of the biggest bookmakers, and it offers low margins) in each season and league. It is clear that the best values of $LogLoss$ function are obtained for the Premier League with very different results from other leagues. Quite surprisingly, the second best results are obtained for the National League (EN_5). The second surprise is that there is no clear trend in time, and it does not seem that models of bookmakers are improved in the view of $LogLoss$ function. The lowest values of $LogLoss$ were usually recorded by $B365, VC$, and WH.

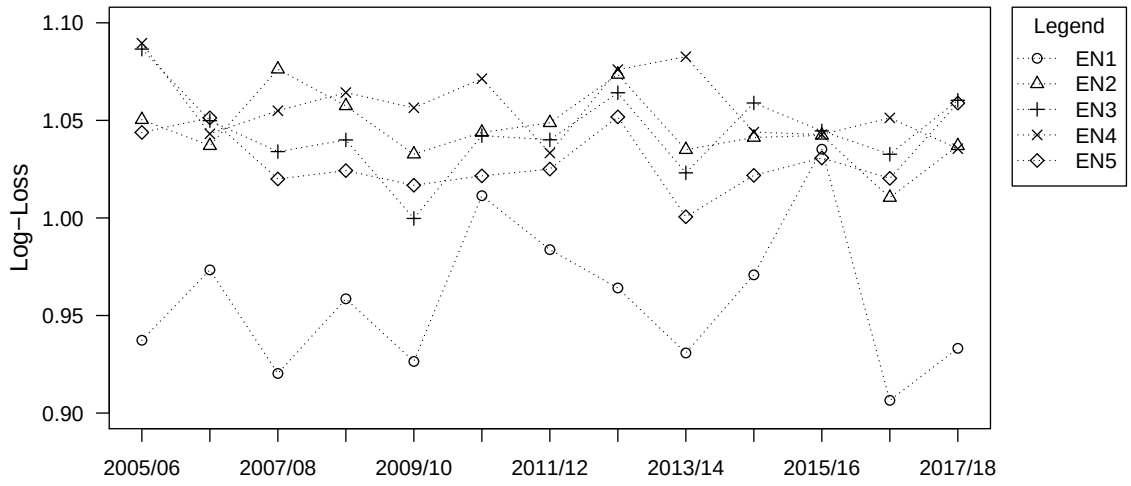


Figure 3 $LogLoss$ function of $B365$

3.4 Prediction Accuracy

The prediction accuracy is the percentage of matches that ended with the result that was the most probable according to the bookmaker. As in the case of calibration, we have to resolve small number of cases, where two or even all three outcomes have the same probability. In this case, both outcomes (or in rare situation all three outcomes) are counted as a correct prediction. This means, that the baseline for prediction accuracy is little higher than 1/3, i.e. random prediction with three possible outcomes and in small number of cases with two or even one possible outcome (e.g., for a situation where 10 matches out of 380 have only two possible outcomes, the baseline is 0.338).

Differences between bookmakers are again relatively small when compared to differences between consecutive seasons. If we exclude results of WH in the 2007/08 season (because of 1/3 of missing odds in this season), then the highest difference was recorded in the 2014/15 season between $B365$ (0.493) and WH (0.533) in EN_4. As stated before, higher differences of accuracy are common between seasons; therefore, Figure 4 presents results only of $B365$ in each season and league. It is clear that the best results are obtained for the Premier League and usually with very different results from other leagues. The second best results are again obtained for the National League (EN_5). As in the previous part, there is no clear trend in time. Results of the prediction accuracy are consistent with the results of the $LogLoss$ function. The comparison of bookmakers is not as clear as in the $LogLoss$ function. However, if we restrict ourselves only on the Premier League, then the best results (almost in each season) are recorded by IW and poor results are recorded by $B365, BW$, and VC. We remind that the lowest values of $LogLoss$ were usually recorded by $B365, VC$, and WH.

4 Conclusion

This paper investigated margins of six well known bookmakers and qualities of their models in five top level English leagues between the 2005/06 season and 2017/18 season. It is clear that margins were lowered in recent years, mainly in the Premier League. Bookmakers' models were compared using three measures – calibration, $LogLoss$ function, and prediction accuracy. These measures were also used to evaluate changes of models in time. Calibration showed that differences among bookmakers are usually lower than differences between two

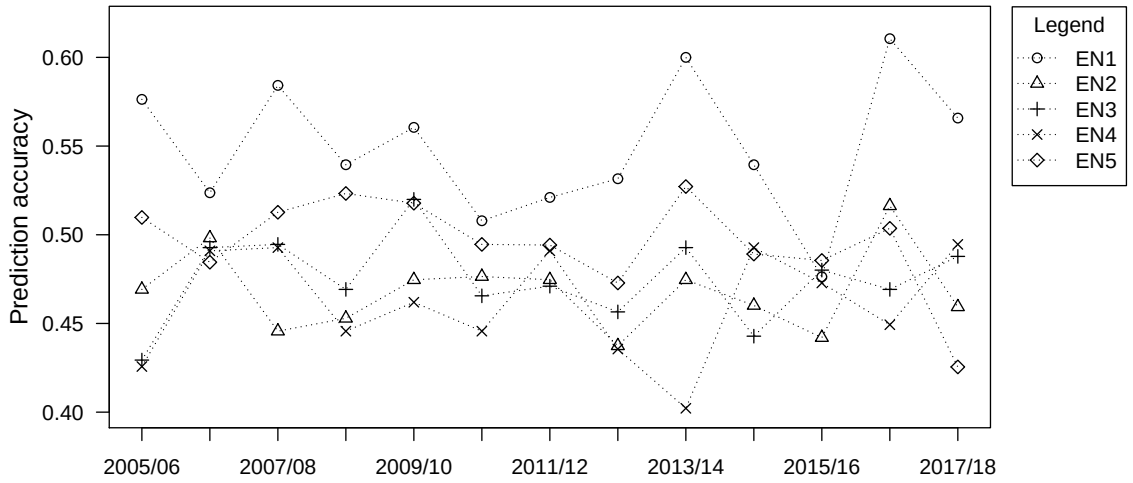


Figure 4 Prediction accuracy of $B365$

consecutive seasons. One possible explanation is that bookmakers use very similar models. This was confirmed by two following measures. Calibration also showed that models can be considered as good; however, no clear improvement in time was seen. From the view of $LogLoss$ function the best results of bookmakers' models were recorded in the Premier League, with other leagues substantially behind. This result was also confirmed by the prediction accuracy of models. There is no clear winner among bookmakers, and their models can be considered as similar. Finally, no significant improvement of models was recorded in the selected seasons.

Acknowledgements

This publication was supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports.

References

[1] Buttrey, S. E. (2016). Beating the market betting on NHL hockey games. *Journal of Quantitative Analysis in Sports*, 12(2), 87–98.

[2] Che, X., Feddersen, A. & Humphreys, B. R. (2017).Price setting and competition in fixed odds betting markets. In P. Rodríguez, B. R. Humphreys & R. Simmons (Ed.). *The Economics of Sports Betting* (pp. 38–51). Cheltenham: Edward Elgar Publishing Limited.

[3] Dixon, M. J. & Coles, S. G. (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 46, 265–280.

[4] Football-Data.co.uk (2018). *Football Results, Statistics & Soccer Betting Odds Data*. (Retrieved March 28, 2018, from http://www.football-data.co.uk/data.php)

[5] Karlis, D. & Ntzoufras, I. (2003). Analysis of Sports Data by using Bivariate Poisson Models. *The Statistician*, 52, 381–393.

[6] Kovalchik S. A. (2016). Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*, 12(3), 127–138.

[7] Maher, M. J. (1982). Modelling Association Football Scores. *Statistica Neerlandica*, 36, 109–118.

[8] Malarić, R., Katić, T. & Sabolić, D. (2008). The market efficiency of the soccer fixed odds internet betting market. *Applied Economics Letters*, 15(3), 171–174.

[9] Marek, P., Šedivá, B. & Ťoupal, T. (2014). Modeling and prediction of ice hockey match results. *Journal of Quantitative Analysis in Sports*, 10(3), 357–365.

[10] Statista.com (2018). *Market value of online gambling worldwide 2009–2020*. (Retrieved April 10, 2018, from https://www.statista.com/statistics/270728/market-volume-of-online-gaming-worldwide/)

[11] Šedivá, B. (2015). Dynamic Forecasting Model Based on Ranking System for Hockey Match. In *Mathsport International 2015 Conference Proceedings.*, Loughborough: Loughborough University, 177–182.

[12] Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30, 934–943.

Comparison of Expected Coalition Influence for Two Survey-Based Parliamentary Settings

Elena Mielcová[1]

Abstract. The main aim of this article is to compare an influence of political parties in two hypothetical distributions of seats in the Chamber of Deputies of the Parliament of the Czech Republic. Both discussed hypothetical seat distributions were obtained in survey conducted in 2016-2017. The first discussed seat distribution is based on the elections of the current electoral system, the second seat distribution is based on the proposed electoral system called "Systém Demokracie 2.1" (also Democracy 2.1) which is taking into account both positive, and negative votes.
For obtained data, the a-priori influence of political parties is measured by power indices. Moreover, the idea of a-priori coalition preferences should be taken into account. As the main aim of this article is to study of coalition formation on data simulating results of election in real voting body, in order to cover uncertain character of real data, it was necessary to add uncertainty issues into the concept of power index values – namely the concept of Atanassov intuitionistic fuzzy (AI-fuzzy) sets approach.

Keywords: Power index, a-priori coalition structure, elections, AI-fuzzy sets.

JEL Classification: C71
AMS Classification: 91A80

1 Introduction

The main aim of this article is to compare an influence of political parties in two hypothetical distributions of seats in the Chamber of Deputies of the Parliament of the Czech Republic. Both discussed hypothetical seat distributions were obtained in survey conducted in 2016-2017 and were presented as the Diploma Project defended at OPF SU in 2017 [6]. While the first discussed hypothetical seat distribution is based on the current electoral system, the second seat distribution is based on the proposed electoral system called "Demokracie 2.1" or "Democracy 2.1", which is taking into account both positive, and negative votes [7].

For obtained data, the a-priori influence of political parties is measured by power indices. Moreover, the idea of a-priori coalition preferences should be taken into account [9]. As the main aim of this article is to study of coalition formation on data simulating results of election in real voting body, in order to cover uncertain character of real data, it was necessary to add uncertainty issues into the concept of power index values – namely the concept of Atanassov intuitionistic fuzzy sets approach [1]. Dubois et al. [5] discussed terminological difficulties concerning term "an intuitionistic fuzzy set" and proposed use of different term; therefore throughout this text, the term "AI-fuzzy" will be used instead of the term "Atanassov intuitionistic fuzzy".

This text is organized as follows: the description of used seats recalculation rules are given in the next part, followed by preliminaries covering weighted games, calculation of Shapley-Shubik power index, and Banzhaf power index. The data description part is followed by calculation results. Conclusions and list of references end the text.

2 Elections to the Chamber of Deputies of the Parliament of the Czech Republic

In the Czech Republic, the Czech Constitution states that deputies of the Chamber of Deputies are elected *"…on the basis of universal, equal, and direct secret ballot voting according to the principles of proportional representation…*". All 200 members of the Chamber of Deputies are elected for a four-year period; the elections run in fourteen regions. Before the elections, each political party (or coalitions of two or more parties) proposes a list of ranked candidates for every region. During elections, voters choose a political party. After elections, the number of seats in the Lower House for every region is determined by the ratio of voters in that region. Votes for the entire republic are then counted, and parties which receive less than 5 % votes (or coalitions of two, three or more parties

[1] Silesian University in Opava, School of Business Administration in Karviná / Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, Karviná, Czech Republic, mielcova@opf.slu.cz.

with less than 10 %, 15 % or 20 % respectively) are eliminated. However, the state covers some cost of every vote gained by a party that won at least 1.5 % of valid votes.

The number of seats parties receive is in proportion to their total number of votes, the D'Hondt method of determination of number of seats is used. The number of regional party representatives depends on the proportion of votes for that party in that region. Representatives are taken from a list of delegates in the preliminary ranking, if the voters did not use the right to choose another order (so-called preference votes). Under these voting rules, election campaigns do not focus on particular people as their representatives but rather on a political party and its political program. Czech voters are choosing a concrete political party first and only secondly the representatives of the region. The description of actual electoral rules can be found at official web site of the Parliament of the Czech Republic.

The electoral rules were slightly modified since 1990s; the biggest changes were in 1990s concerning number of regions, and a change in threshold for coalitions of several political parties. There are also critiques of this electoral system; the most popular of critics is the Czech businessman Karel Janeček, who proposed his new electoral system, called Democracy 2.1.

Democracy 2.1 method [7] is a semi-proportional system, expecting more electoral regions – districts. In each district, two representatives are elected. Hence any competing party nominates only one or two candidates; moreover, independent candidates may compete subject to other conditions (for example, upon gathering a certain number of signatures from eligible voters). Voters may cast up to four plus-votes and up to two minus-votes which they may distribute across all candidates in the district. For every voter, one minus vote is valid if at least two plus votes were cast, and, similarly, two minus votes are valid if four votes were cast. In the case of lesser number of plus votes, only plus votes are summed up. A voter may give only one vote to any candidate.

Janeček [7] discusses possibilities of manipulation as well as suitability of the proposed method. He argues, that "…*We expect that in a parliament or legislative body elected under the D21system, there will be more smaller parties represented than under a typical proportional system. At the same time, a strong party with broad electoral appeal might become even stronger, thanks to the "majoritarian" effect of our system, than it would under a proportional system. What kind of party will weaken? Medium-size parties, especially those with more narrow electoral appeal, are likely to win fewer seats under D21 than under a typical proportional system*…" Moreover, he expects easier formation of government which should become more stable. The system expects smaller electoral regions copying Senate electoral regions; as number of these regions is 81, and there are two expected deputies from each region, in accord with the proposed Democracy 2.1 electoral system the House of Deputies should shrink down to 162 seats.

3 Preliminaries

Parliamentary voting can be described as a simple weighted game: let N be a set of all players, let $S \subset N$ be a coalition of players from the set of all players. A characteristic function game is a pair $[N,v]$ consisting of a set of players $N = \{1,2,\dots,n\}$ and a characteristic function v which maps every $S \subset N$ to a nonnegative number $v(S)$ with condition $v(\emptyset) = 0$. In the simple game, for every coalition $S \subset N$, $v(S) = 1$ if S wins and $v(S) = 0$ if S loses. The Czech Parliamentary voting can be described as the special type of a simple game – the weighted voting game – in which political parties are players of the game with different weights. Let $N = \{1,2,\dots,n\}$ be a set of n players, let $w = (w_1, w_2, \dots, w_n)$ be a vector of players' weights, and let q be a quota. Then the triple $G = [N, w, q]$ is called a weighted voting game. A coalition S is winning if its total weight meets or exceeds the quota q, that means:

$$\sum_{i \in S} w_i \geq q \tag{1}$$

The Shapley–Shubik power index [11] was created to a-priori evaluation of the power division among bodies in committee system. The derivation of the member's Shapley-Shubik power index is based on the Shapley value [10]:

$$\varphi_p = \sum_{p \in T \subset N} \frac{(t-1)!(n-t)!}{n!}(v(T) - v(T \setminus p)) \tag{2}$$

The Shapley value in the case of simple games:

$$\varphi_p = \sum_{\substack{p \in T \subset N \\ T \text{ winning} \\ T \setminus p \text{ losing}}} \frac{(t-1)!(n-t)!}{n!} \tag{3}$$

where summation is done through all winning coalitions $T \subset N$ containing player p such that a coalition that is created from T by omission of player p (denoted $T\backslash p$) is losing. The cardinality of the set T is denoted by t.

The idea behind the Banzhaf power index is based on calculations of swing players – players that are crucial for a coalition to be winning. Let player p be in a coalition T. For any $T \subset N$ we say player p swings in T if coalition T is winning and coalition $T\backslash p$ is losing. If p is a swing player of a coalition T, then $v(T) - v(T\backslash i) = 1$. The originally proposed Banzhaf power index [2,4] counts the number of swings over all possible nonempty coalitions; the normalized Banzhaf index [2] calculates voter p's swings to total amount of swings.

$$b_p = \sum_{\substack{p \in T \subset N \\ T \text{ winning} \\ T\backslash p \text{ losing}}} \frac{1}{2^{n-1}} \tag{4}$$

The normalized Banzhaf index for voter p is obtained by dividing the sum of p's swings (regarding all possible combinations) by the sum of all voters' all swings. Formally voter p's standardized Banzhaf index is calculated as:

$$b_p^n = \frac{\theta_p}{\theta} \tag{5}$$

where θ_p denotes the number of voter p's swings, and θ is a total number of swings of all players.

An AI-fuzzy set is defined as a set of triples $A = \{\langle x_i, \mu_A(x_i), \nu_A(x_i)\rangle\}$ over a fixed set $X = \{x_1, x_2, \dots x_n\}$. Functions $\mu_A: x \to L$ and $\nu_A: x \to L$ for L=[0,1] define the degree of membership and the degree of non-membership of the element $x_i \in X$ to $A \subseteq X$. For an AI-fuzzy set, the condition $0 \leq \mu_A(x_i) + \nu_A(x_i) \leq 1$ holds for all $x_i \in X$ [1].

In this analysis, weights of players will be expressed as AI-fuzzy numbers, defined in [3]. Let T be a crisp coalition such that $T \subset N$. Let $\tilde{C}$ be an AI-fuzzy coalition. Let e^T be an AI-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 $T \subset N$ for the referral coalition $\tilde{C}$ can be expressed as:

$$\Phi_i(v)_{\tilde{S}} = \sum_{i \in T \subset N} \frac{(t-1)!(n-t)!}{n!} \left(v\left(\sum_{j \in T} \tilde{S}_j e_j^T\right) - v\left(\sum_{j \in T\backslash i} \tilde{S}_j e_j^T\right)\right) \tag{6}$$

In the case of Banzhaf value, possible swings are calculated with respect to respective membership and nonmembership functions. In this case there the referral coalition $\tilde{C}$ is considered to be "a typical coalition" for a played game (for example in voting game it can be a 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 Parliament of the Czech Republic, a probabilistic approach was used, characteristic function values were calculated such that (see also [8]):

- if the sum of participation of party members in coalition T (= minimal(T)) is greater or equal than quota, then v(T)=1 (for example in the current Chamber of Deputies of the Parliament of the Czech Republic, if all 200 members are present, then the simple majority rule expects that more than 100 votes ensure passing a bill, hence the quota is 101)
- if the sum of participation of party members in coalition T and all values of uncertainty of participation for all political parties (= maximal(T)) is smaller than quota, then v(T)=0.
- if the sum of participation of party members in coalition T is smaller than quota, and the sum of participation of party members in coalition T and all values of uncertainty of participation for all political parties is greater or equal than quota, then the respective value v(T) is calculating as the value of v(T) = 1 - F(quota), where F(quota) is the value of a cumulative distribution function for the continuous uniform distribution (unif(minimal(T); maximal(T))).

4 Data Description

This analysis is based on the data from the survey conducted in winter 2016/2017 as a part of diploma project at OPF SU Karvina [6]. Respondents were given the choice based on political parties present in 2013 Elections. All of them received choice of political party for the D'Hondt seat distribution system, as well as description of Democracy 2.1 seats distribution system with list of two leading political members in respective region (from original

electoral lists). The survey was filled in by 335 respondents from the whole Czech Republic – the number is quite small to explain political atmosphere in the country, however it can sketch the situation in decision making process in this specific group of people. The age distribution of respondents – 29% of respondents from group 18-30 years old, 44% from the age group 31-50 years old, and 27% more than 51 years old.

Distribution of seats in a hypothetical parliament based on a conducted survey is presented in Figures 1 and 2 in the case of original Czech electoral system, as well as Democracy 2.1 electoral system, respectively. While the current system has distributed 200 seats among 9 political parties, the Democracy 2.1 system has distributed 162 seats among 3 political parties. Moreover, with respect to survey results, the second strongest political party when considering use of current system (ANO political party) would not get any seat considering Democracy 2.1 electoral system.

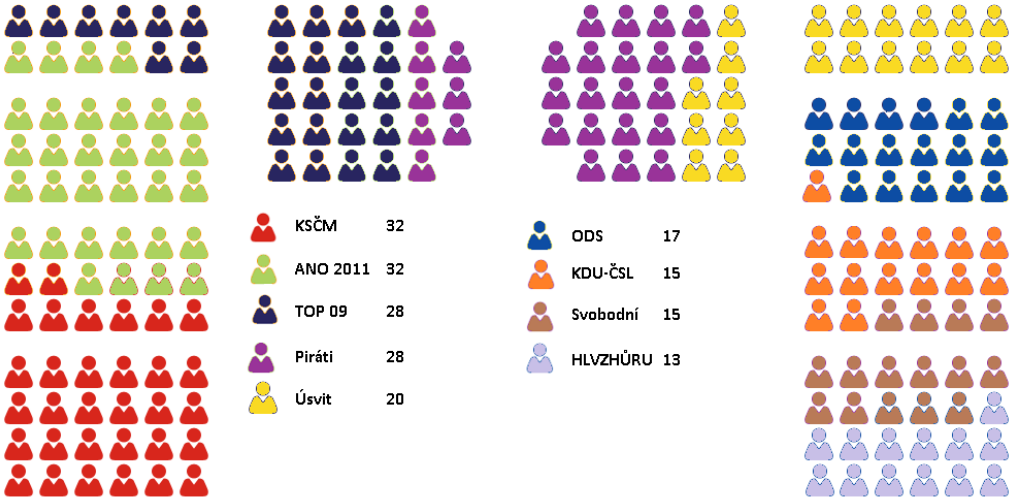


Figure 1 Distribution of seats in the hypothetical Chamber of Deputies with respect to survey data and current electoral system. *Source: Fusková, M. (2017). Diploma Project OPF SU Karviná [6]*

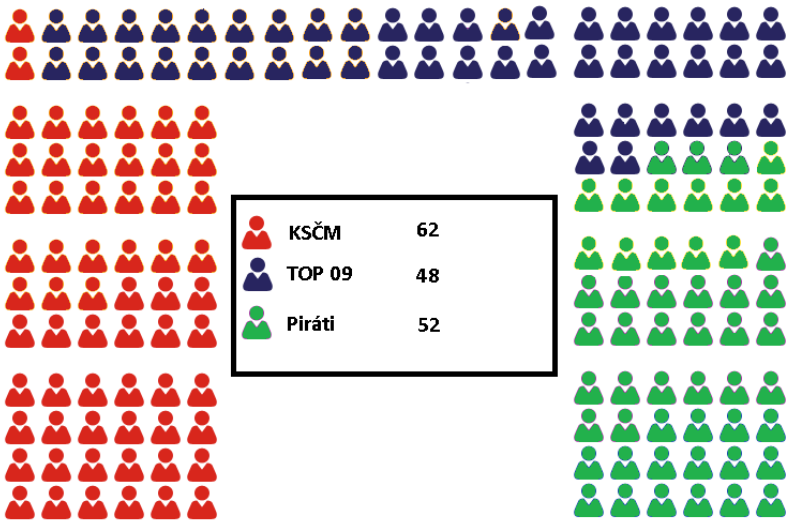


Figure 2 Distribution of seats in the hypothetical Chamber of Deputies with respect to survey data and Democracy 2.1 electoral system. *Source: Fusková, M. (2017). Diploma Project OPF SU Karviná [6]*

In order to introduce a little bit more "reality" into voting power calculations, the AI-fuzzy approach was used. The participation and nonparticipation levels of a referral coalition $\tilde{C}$ were estimated from real parliamentary voting data. Table 1 shows membership and nonmembership levels of "a typical coalition" calculated as average participation of political parties in winning coalitions and respective participations in losing coalitions for 2006-2013 Chamber of Deputies [8]. In calculations of AI-fuzzy indices, the membership and nonmembership levels of

political parties as in Table 1 were used. In the case, when political party was not present in 2006-2013 Chamber of Deputies of the Parliament of the Czech Republic, the average value of membership and nonmembership levels were considered as political party appropriate levels (values in last column of Table 1).

Political party 2010-2013	ČSSD	ODS	TOP09	KSČM	VV	average
Membership level	0.392	0.672	0.761	0.453	0.592	0.574
Nonmembership level	0.354	0.087	0.072	0.407	0.156	0.215

Table 1 Typical party membership and nonmembership levels for 2006-2013 Parliaments. *Source: [8].*

5 Results

The standard way how to measure power of players – in our case political parties – is to evaluate them by predefined indices. All results of calculations of the a-priori political power evaluations of political parties in two hypothetical parliaments based on current voting system, as well as based on Democracy 2.1 voting system are given in Tables 2 and 3, respectively.

When considering current voting system, the parliament, according to survey, should be composed of nine political parties. Taking into account a-priori evaluation by standard Shapley-Shubik and Standardized Banzhaf power indices (in Table 2), four strongest political parties would share more than 60% of whole decision power in this type of parliament. The power is distributed among these four political parties approximately evenly – first two political parties (KSČM and ANO 2011) have slightly higher power index than second two political parties (TOP 09 and Piráti). It is not easy to estimate potential governmental coalition; hence, no power index with predefined coalition structure was calculated. However, from numerous political parties' statements and negotiations present after 2017 elections we could guess that no coalition would be created with the Communist Party (KSČM). Even without the guess, from the calculated AI-fuzzy indices that taking into account values from previous parliamentary period (see Table 1), the power index of KSČM is significantly smaller comparing to other three strongest political parties.

Political party	Seats	Shapley-Shubik	Standard. Banzhaf	AI-fuzzy Shapley-Shubik	AI-fuzzy Banzhaf
KSČM	32	0.163	0.160	0.122	0.119
ANO 2011	32	0.163	0.160	0.163	0.149
TOP 09	28	0.156	0.154	0.185	0.157
Piráti	28	0.156	0.154	0.137	0.132
Úsvit	20	0.077	0.078	0.096	0.096
ODS	17	0.077	0.078	0.095	0.095
KDU-ČSL	15	0.070	0.072	0.071	0.072
Svobodní	15	0.070	0.072	0.071	0.072
Hlavu vzhůru	13	0.070	0.072	0.060	0.063

Table 2 Typical party participation and nonparticipation levels for 2006-2013 Parliaments. *Source: own calculations.*

When considering Democracy 2.1 voting system, according to survey the parliament should be composed of three political parties (see Table 3). The interesting point is the fact that the second strongest political party under current voting system (ANO 2011) would not be in this parliament because of the negative votes.

Political party	Seats	Shapley-Shubik	Standard. Banzhaf	AI-fuzzy Shapley-Shubik	AI-fuzzy Banzhaf
KSČM	62	0.333	0.333	0.2744	0.2058
TOP 09	48	0.333	0.333	0.3628	0.2942
Piráti	52	0.333	0.333	0.3628	0.2942

Table 3 Typical party participation and nonparticipation levels for 2006-2013 Parliaments. *Source: own calculations.*

Taking into account a-priori evaluation by standard Shapley-Shubik and Standardized Banzhaf power indices, all political parties would share power evenly. Taking into account expected coalition structure TOP 09 + Piráti

(no party is in coalition with KSČM), then the power index distribution would be 0 for KSČM, 0.5 for TOP 09 and 0.5 for Piráti political party. As in the previous parliament distributions, from the calculated AI-fuzzy values of Shapley-Shubik power index and Banzhaf power index (without normalization) we can see that the power index of KSČM is significantly smaller comparing to power index of other two political parties.

6 Conclusion

The main aim of this article was to compare the potential influence of political parties in two hypothetical distributions of seats in the Chamber of Deputies of the Parliament of the Czech Republic. Both discussed hypothetical seat distributions were obtained in survey conducted in 2016-2017, the first distribution is calculated with respect to current electoral system, while the second one is calculated with respect to Democracy 2.1 electoral system proposed by Czech businessman Karel Janeček.

Even though the original survey was conducted only on small sample of respondents, the seats distribution under these two electoral rules are substantially different mainly in number of political parties present in both types of parliament. The interesting point is the fact that the second strongest political party under current voting system (ANO 2011) would not be in Democracy 2.1 type of parliament because of the negative votes. When comparing classical a-priori power indices of political parties in both parliaments (namely Shapley-Shubik power index and Banzhaf power index) with calculated AI-fuzzy a-priori power indices, there is a significant decrease in power of the Communist party. This result corresponds with government negotiations after 2017 elections, when most of political parties declared that they are not making coalition with the Communist Party – KSČM (see for example comments in [12,13]).

Acknowledgements

Supported by the grant No. 18-01246S of the Grant Agency GAČR.

References

[1] Atanassov, K. T. (1986). Intuitionistic Fuzzy Sets. *Fuzzy Sets and Systems, 20(1)*, 87–96.
[2] Banzhaf III, J. F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review, 19*, 317-343.
[3] Burillo, P., Bustince, H. & Mohedano, V. (1994). Some definitions of intuitionistic fuzzy number. First properties. In: *Proceedings of the 1st Workshop on Fuzzy Based Expert Systems*, 53-55.
[4] Coleman, J. S. (1971) Control of Collectivities and the Power of a Collectivity to Act. In B.Lieberman (Ed.), *Social Choice*, Gordon and Breach, London, New York.
[5] Dubois, D., Gottwals, S., Hajek, P., Kacprzyk, J. & Prade, H. (2005). Terminological difficulties in fuzzy set theory – the case of 'intuitionistic fuzzy sets', *Fuzzy Sets and Systems, 156*, 485-491.
[6] Fusková, M. (2017). *Systém demokracie 2.1 a volby do PSP ČR 2013.* Diplomová práce OPF SU Karviná.
[7] Janeček, K. (2016). Democracy 2.1. Available at: https://uploads-ssl.webflow.com/5a61da42ae3f8f00019b46bd/5a893542d9716c0001a4a80d_d21.pdf [cit. 2018-04-18]
[8] Mielcová, E. (2017). Expected Coalition Influence under I-Fuzzy Setting: The Case of the Czech Parliament. In: Pražák P. (Ed.), *35th International Conference Mathematical Methods in Economics. Conference Proceedings* (pp. 448-453) Hradec Králové: University of Hradec Králové.
[9] Owen, G. (1977). Values of Games with a priori Unions. In: R. Henn & O. Moeschlin (Eds.), *Essays in mathematical economics and game theory* (pp.76-88). Berlin: Fields Institute.
[10] Shapley, L. S. (1953) A value for n-Person Games. In: H. W. Kuhn & A. W. Tucker (Eds.*): Contribution to the Theory of Games 2* (pp. 307-317). Princeton, New Jersey: Princeton University Press.
[11] Shapley, L. S. & Shubik, M. (1954) A Method for Evaluating the Distribution of Power in a Committee Systém. *American Political Science Review, 48*, 784–792.
[12] Web: eurozpravy.cz. (2017) Babiš nevyloučil vládu s ČSSD, odmítl ji ale s KSČM a Okamurou. Kalousek prsty v Čapím hnízdě nemá, míní. (In English: Babiš did not rule out the government with the CSSD, but rejected it with KSČM and Okamura. Kalousek has nothing to do with "Čapí hnízdo", he thinks.) Available at: http://domaci.eurozpravy.cz/politika/202936-babis-pripousti-opozici-do-vlady-s-kscm-a-okamurou-ne-pujde-zeman-at-vymeni-personal-vzkazal/ [cit. 2018-04-28]
[13] Web: idnes.cz. (2017) ANALÝZA: Do koalice s ANO, KSČM nebo TOP 09 se nikdo moc nehrne. (In English: ANALYSIS: Nobody wants to be in a coalition with ANO, KSČM or TOP 09.) Avilable at: https://zpravy.idnes.cz/parlament-snemovna-volby-2017-ano-ods-cssd-top09-lidovci-pd1-/domaci.aspx?c=A171003_130714_domaci_jav [cit. 2018-04-28]

The TMAI measure versus chosen measures of deterministic chaos. The comparison of investment strategies

Monika Miśkiewicz-Nawrocka[1]

Abstract. In the 1950s H. Markowitz started intensive development of scientific field which is a portfolio analysis. It is one of the most important techniques for investing in the capital market, which main goal is to diversify the investment risk. In recent years, 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, taxonomic measure of investment attractiveness (TMAI), value investing, growth investing, dividend investing, innovation investing, csr investing. 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. Studies carried out so far had shown that the use of above-mentioned measures has a significant impact on the construction of the optimal portfolio.

The paper aims to construct optimal portfolios determined based on the largest Lyapunov exponent, the Hurst exponent and the TMAI measure and next to evaluate and compare their efficiency. The test will be conducted on the basis of stocks prices time series of companies from WIG20 index.

Keywords: portfolio analysis, TMAI measure, largest Lyapunov exponent, Hurst exponent.

JEL Classification: C3, C8, G11, E4
AMS Classification: 91B28

1 Introduction

The intensive development of the portfolio analysis observed in the last half-century has led to the creation of many new methods to estimate the composition of the optimal portfolio. 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) [13,14], value investing, growth investing, dividend investing [2, 17], innovation investing . A new approach [8,9] 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. Studies carried out so far had shown that the use of the above-mentioned measures gives good results [8,9,10,11].

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, the Hurst exponent and the taxonomic measure of investment attractiveness TMAI. 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 1.01.1999 to 31.12.2016, and stocks portfolios were built at the end of each year between 2008 and 2015.

2 The TMAI measure

The taxonomic measure of investment attractiveness (TMAI) is one of the methods for the selection of companies [13, 16]. This method allows for a comprehensive evaluation of the companies based on key financial and market indicators and present them in the form of synthetic measure.

Building the taxonomic measure consists of three stages [7, 14, 15]. First, we create 2-dimensional data matrix containing observations x_{ij} of the diagnostic features of selected objects. We normalize (standardize) these values, following the formula:

[1] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, monika.miskiewicz@ue.katowice.pl.

$$y_{ij} = \left(x_{ij} - \bar{x}_j\right) / S_j, \quad i = 1,\ldots,n; \quad j = 1,\ldots,m, \tag{1}$$

where: x_{ij} - value of *j*-th diagnostic features for *i*-th object, $\bar{x}_j$ mean of feature *j*, S_j standard deviation for *j*. Next, the module method is used, and in the normalized matrix of *m* variables, the highest value is taken, module y_{0j}. The Euclidean distance from the module is calculated, using the formula:

$$d_i = \left[\frac{1}{m}\sum_{j=1}^{m}\left(y_{ij} - y_{0j}\right)^2\right]^{\frac{1}{2}}, \quad i = 1,\ldots,n. \tag{2}$$

The shorter the distance of the given object from the module, the lower is the value d_i. The obtained variable is not normalized, which next is transformed into a stimulant using the formula:

$$TMAI_i = 1 - \frac{d_i}{d_0}, \quad i = 1,\ldots,n, \tag{3}$$

where: $TMAI_i$ – taxonomic development measure for object *i*, d_i – distance of *i* object from module, d_0 – standard to assure that variable $TMAI_i$ will take values ranging from 0 to 1, for example

$$d_0 = \bar{d} + 2S_d,$$

where: $\bar{d}, S_d$ - mean and standard deviation d_i.

3 The largest Lyapunov exponent

For the discrete dynamic system [5] described by first order recurrence equation:

$$x_{t+1} = f\left(x_t\right), \qquad t = 0,1,2,\ldots, \tag{4}$$

where x_t, x_{t+1} - state on the system respectively at the moments *t* and $t+1$, the Lyapunov exponents are defined as limits [18]

$$\lambda_i\left(x_0\right) = \lim_{n\to\infty}\frac{1}{n}\ln\left|\mu_i\left(n, x_0\right)\right|, \qquad i = 1, ..., m, \quad \text{for } m \geq 1, \tag{5}$$

where $\mu_i\left(n, x_0\right)$ 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. There are *m* Lyapunov exponents for the *m*-dimensional system. The largest (maximal) of them 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} [3,6].

For real-time series, if you do not know a generator function *f*, the largest (maximal) Lyapunov exponent is estimated based on the relation [18]:

$$\Delta_n = \Delta_0 \cdot e^{n\lambda_{\max}}, \tag{6}$$

as the direction component of the regression equation [4, 5, 9]:

$$\ln\Delta_n = \ln\Delta_0 + \lambda_{\max} n, \tag{7}$$

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 $\left(x_1, x_2, \ldots, x_N\right)$. 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]:

$$\Delta_1 \approx \Delta_{\min} \cdot e^{\lambda_{\max}}, \tag{8}$$

where $\lambda_{\max}$ is the largest (maximal) Lyapunov exponent.

4 The 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 $\{x_1, x_2, ..., x_N\}$ it runs through the following steps[1]:
Firstly, transform the above time series into $m = N - 1$ logarithmic rates of return:

$$y_k = \log(x_{k+1}/x_k), \; k = 1, 2, \ldots, N-1, \tag{9}$$

and share the series (9) 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 $k = 1, 2, \ldots, tT$.
In the next step, define the:

$$z_{ij} = y_{ij} - \bar{y}_j, \tag{10}$$

where: y_{ij} is the j-th value in the i-th interval and $\bar{y}_j = \frac{1}{t}\sum_{i=1}^{t} y_{ij}$.
The range of the i-th interval is defined as:

$$R_j = \max(q_{ij}) - \min(q_{ij}), \tag{11}$$

where: $q_{ij} = \sum_{l=1}^{i} z_{lj}$, $i = 1, 2, \ldots, t$, $j = 1, 2, \ldots, T$ is a sequence of partial sums z_{lj} for each i.
Next, calculate the rescaled range series (R/S):

$$(R/S)_t = (1/T)\sum_{j=1}^{T} \alpha_{jt}. \tag{12}$$

$$\alpha_{jt} = R_j / S_j, \tag{13}$$

where: $S_j = \sqrt{\frac{1}{t}\sum_{i=1}^{t} z_{ij}^2}$.

The above procedure is carried out for different lengths of time series t. 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.

5 The optimal stock portfolio

The level of chaos, system persistence or investment attractiveness are treated as positive phenomena and therefore the following maximization problems are solved in order to determine the optimal allocation of shares in the portfolio [8, 9]. The higher the value of TMAI, the greater the attractiveness of the company. The greater the value of Lyapunov exponent, the greater predictability of the system. The greater the value of the Hurst exponent, the greater the system's persistence (i.e. reinforcing trend).

Model I	**Model II**	**Model III**
$\max\left(\sum_{i=1}^{m} \lambda_{\max i} x_i\right)$	$\max\left(\sum_{i=1}^{m} H_i x_i\right)$	$\max\left(\sum_{i=1}^{m} TMAI_i x_i\right)$
$R_p \geq R_0$	$R_p \geq R_0$	$R_p \geq R_0$
$\sum_{i=1}^{m} S_i x_i \leq S_0$	$\sum_{i=1}^{m} S_i x_i \leq S_0$	$\sum_{i=1}^{m} S_i x_i \leq S_0$
$\sum_{i=1}^{m} x_i = 1$	$\sum_{i=1}^{m} x_i = 1$	$\sum_{i=1}^{m} x_i = 1$
$0 \leq x_i \leq 0.4, \quad i = 1,\ldots,m$	$0 \leq x_i \leq 0.4, \quad i = 1,\ldots,m$	$0 \leq x_i \leq 0.4, \quad i = 1,\ldots,m$

Table 1 Optimization problems

where: S_p - the risk of the portfolio *m*-shares: $S_p^2 = \sum_{i=1}^{m} x_i^2 S_i^2 + 2\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} x_i x_j S_i S_j \rho_{ij}$,

S_i - the standard deviation for *i*-company,

ρ_{ij} - the correlation coefficient of *i*-share with *j*-share,

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,

$TMAI_i$ - the taxonomic measure of investment attractiveness for *i*-company,

$\lambda_{\max i}$ - the largest (maximal) Lyapunov exponent for *i*-company,

H_i - the Hurst exponent for *i*-company.

6 The purpose and conduct of the study

In the study I 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 2008 to 2015. The optimal portfolios were built at the end of each year between 2008-2015.

The value of taxonomic measure were estimated on the basis of the algorithm described in point 1. I used data contained in financial reports of companies [11] for the third quarter of 2013[2]. The largest Lyapunov exponent and the Hurst exponent for the analyzed companies were estimated on the basis of the algorithms described in points 2 and 3. 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 1.01.1999 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 - model III. 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.

Model I	**Portfolio 2008**	**Portfolio 2009**	**Portfolio 2010**	**Portfolio 2011**	**Portfolio 2012**	**Portfolio 2013**	**Portfolio 2014**	**Portfolio 2015**
Rate of return	0.002983	-0.000484	-0.000240	0.000853	-0.000133	-0.000169	-0.000076	0.000231
Risk	0.000357	0.000384	0.000118	0.000171	0.000117	0.000109	0.000099	0.000108
Model III	**Portfolio 2008**	**Portfolio 2009**	**Portfolio 2010**	**Portfolio 2011**	**Portfolio 2012**	**Portfolio 2013**	**Portfolio 2014**	**Portfolio 2015**
Rate of return	0.003393	-0.000802	0.110813	0.000910	-0.000476	0.000027	0.000187	0.400357
Risk	0.000278	0.000262	0.000094	0.000205	0.000115	0.000160	0.000140	0.000138
Model II	**Portfolio 2008**	**Portfolio 2009**	**Portfolio 2010**	**Portfolio 2011**	**Portfolio 2012**	**Portfolio 2013**	**Portfolio 2014**	**Portfolio 2015**
Rate of return	0.004234	-0.000802	0.110813	0.001017	-0.000816	-0.000181	0.000121	0.333798
Risk	0.000366	0.000445	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

On the basis of the data presented in Table 1 we can conclude that for the optimization model II and model III the almost all 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 and model III portfolios in 2010 and 2015 at one of the lowest risk level. Moreover, the higest level of risk was obtained for the optimization model II portfolios in each years of period 2008-2015.

[2] The data come from the author's own calculations based on the financial reports of companies.

[3] The data comes from stooq.com. Date of access: 12.29.2009, 2.12.2010, 10.30.2012, 4.24.2017.

Table 2 shows the obtained annual rates of return for the designated portfolios and for the optimization models.

Model	Portfolio 2008	Portfolio 2009	Portfolio 2010	Portfolio 2011
I	0.401132	0.280969	-0.264658	0.472292
II	0.464493	0.212249	0.014849	0.303723
III	0.572081	0.128595	-0.104365	0.537212
Model	**Portfolio 2012**	**Portfolio 2013**	**Portfolio 2014**	**Portfolio 2015**
I	0.490827	0.050649	-0.242797	0.104305
II	0.492476	-0.001390	-0.341304	0.143249
III	0.320458	0.064478	-0.228618	0.146190

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 2008 and portfolio 2011 in model III, and then in portfolio 2012 in model II and model I. The lowest rate of return obtained for all optimization models in 2014.

Figure 1 compares the rates of return of obtained portfolios with the return of the WIG 20 index for the years 2008-2015.

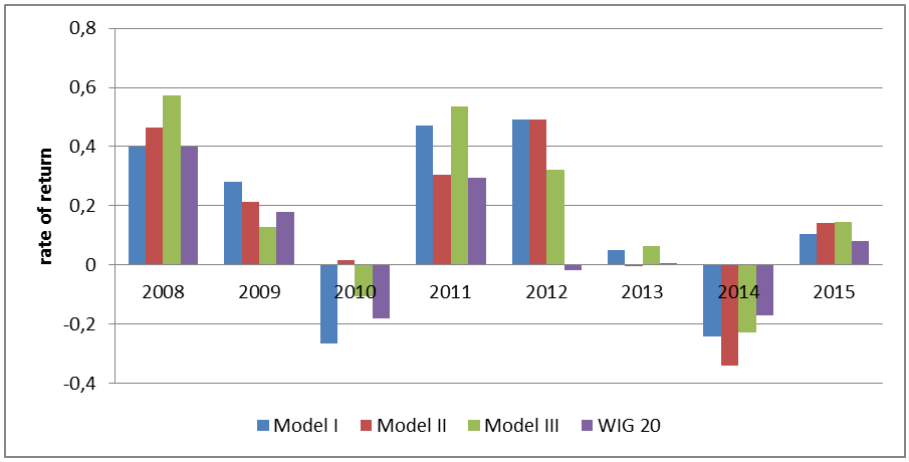


Figure 1 The rate of return of portfolios and index WIG 20

On the basis of the data in Fig. 1 we can conclude that most portfolios in the optimization model I - model III are characterized by higher rate of return than index WIG 20. The exception are years 2010 and 2014. The calculated 8-year return rates for the analyzed strategies showed that the most profitable strategy was to invest in the model III (2.0358), then the II model (1.6363) and model I (1.5450). The worst long-term strategy was WIG 20 (0.5498).

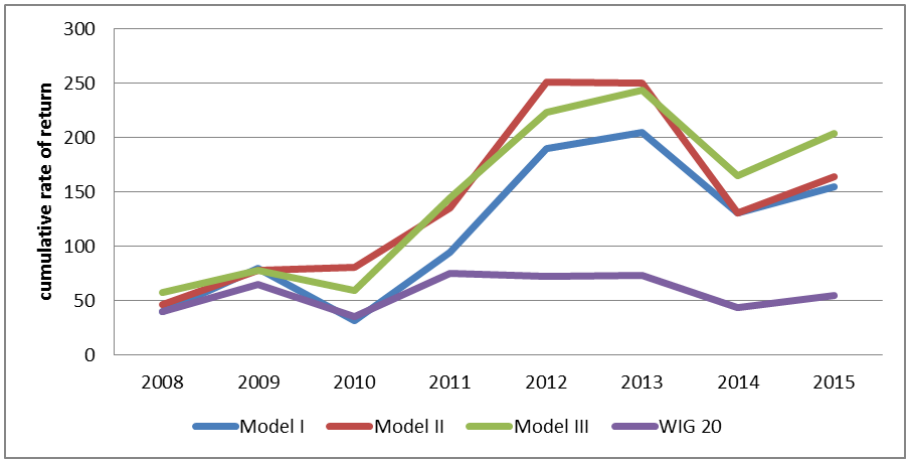


Figure 2 The cumulative rate of return of portfolios and index WIG20

Figure 2 presents the cumulative rate of return of portfolios and index WIG20. Obtained results (fig. 2) show also that the portfolios built on the basis of the model I - model III gives generally better results than index WIG 20.

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, the Hurst exponent and taxonomic measure TMAI. 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 in 2008-2015. Moreover, proposed in the paper approach gives generally better results than the index WIG20 during the considered period.

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.

[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] Kantz, H.: A robust method to estimate the maximal Lyapunov exponent of a time series. *Physical Letters A*, vol. 185(1), 1994, 77 – 87.

[7] 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.

[8] 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.

[9] 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.

[10] Miśkiewicz-Nawrocka M., Zeug-Żebro K.: The efficiency of stocks investment strategy with the use of chosen measures of deterministic chaos to building optimal portfolios, [In:] P. Prazak (ed.) *Proceedings of the 35th International Conference Mathematical Methods in Economics*, University of Hradec Kralove, Hradec Kralove, Czech Republic, 2017, 466-471.

[11] Miśkiewicz-Nawrocka M., Zeug-Żebro K.: The evaluation of the effectiveness of a long-term stocks investment strategy based on the largest Lyapunov exponent , [In:] M. Culik (ed.) *Proceedings of the 11th International Scientific Conference: Financial Management of Firms and Financial Institutions*. VSB - Technical University of Ostrava, Ostrava, Czech Republic, 2017, 590-598.

[12] Nawrocki, T.: Efektywność strategii inwestycji w akcje spółek innowacyjnych notowanych na GPW w Warszawie, *Przegląd Organizacji*, nr 2/2016, s. 53-60.

[13] Tarczyński, W.: A taxonomic measure of the attractiveness of investments in securities. *Przegląd Statystyczny* No 3/94, 1994, 275–300.

[14] Tarczyński, W.: *Fundamentalny portfel papierów wartościowych*. Polskie Wydawnictwo Ekonomiczne, Warszawa, 2002.

[15] Tarczyński, W.: Ocena efektywności metod analizy portfelowej na Giełdzie Papierów Wartościowych w Warszawie za lata 2001-2013. *Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 761, Finanse, rynki finansowe, ubezpieczenia* nr 60, Szczecin, 2013, 537-550.

[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.

Possible and universal robustness of special classes of matrices with inexact data

M. Molnárová[1]

Abstract. Robustness of interval matrices over fuzzy algebra is studied. Fuzzy algebra called also max-min algebra is an extremal algebra with operations maximum and minimum. An interval matrix A over fuzzy algebra (a matrix with inexact data) is a set of matrices given by a lower bound matrix and an upper bound matrix. In this paper special classes of matrices, so called gamma-delta, gamma-delta-row and gamma-delta-column matrices were introduced. Polynomial algorithms for determination of the type and checking the Monge property of above mentioned matrices were proved. Necessary and sufficient conditions for verifying the robustness for matrices with exact data were found. Related polynomial algorithms were described. Sufficient condition for possible robustness and equivalent conditions for universal robustness for matrices with inexact data as well, 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

Extremal algebras (max-min algebra, max-plus algebra, ...) turn modelling in many diverse areas as discrete dynamic systems (DDS), graph theory, knowledge engineering or description of technical devices to rather easy solvable problem [3]. The Monge matrices and their applications were studied in [1], [4]. Robust matrices over fuzzy algebra were investigated in [12]. Sufficient and necessary conditions for robustness of Monge fuzzy matrices were proved in [6]. Robustness of interval fuzzy matrices was studied in [10], [12], [9], [7], [8].

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 [12] to study robustness of a matrix.

Lemma 1. [12] Let $A = (a_{ij}) \in B(n,n)$. Then A is robust if and only if $\mathrm{per}\, A = 1$.

3 Robustness of $\gamma\delta$, $\gamma\delta$-row and $\gamma\delta$-column matrices with fixed data

In this section we introduce special types of matrices, namely $\gamma\delta$, $\gamma\delta$-row and $\gamma\delta$-column matrices. We show that the computational complexity of verifying the type of a matrix and checking the Monge property for special types of matrices is $O(n^2)$. We recall the equivalent conditions for a Monge matrix to be robust. Afterwards we prove necessary and sufficient conditions for $\gamma\delta$, $\gamma\delta$-row and $\gamma\delta$-column matrices 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. However, relevant results regarding concave Monge matrices aren't a trivial consequence of those described in this article.

Equivalent conditions for robustness of Monge matrices in general case of max-min algebra were proved.

Theorem 1. [6] *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 component contains a loop.*

Now, we introduce the matrix of type $\gamma\delta$, $\gamma\delta$-row and $\gamma\delta$-column.

Definition 3. Let $A \in B(m,n)$. Let $\gamma \in \{1,2,\dots,m-1\}$ and $\delta \in \{1,2,\dots,n-1\}$. We say that the matrix A is
a) of type $\gamma\delta$, if

$$a_{ij} = \begin{cases} \text{arbitrary} & \text{if } \ i \in \{\gamma,\, \gamma+1\},\ j \in \{\delta,\, \delta+1\} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

b) of type $\gamma\delta$-row, if

$$a_{ij} = \begin{cases} \text{arbitrary} & \text{if } \ i \in \{\gamma, \gamma+1\}, \delta \leq j \leq n, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

c) of type $\gamma\delta$-column, if

$$a_{ij} = \begin{cases} \text{arbitrary} & \text{if } \ j \in \{\delta, \delta+1\}, \gamma \leq i \leq m, \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Remark 1. There are cases, when the given matrix $A \in B(m,n)$ can be considered as some type of above defined special matrices. First, it is of type $\gamma\delta$, if $a_{ij} = 0$ for all $i \in M,\ j \in N$ ($\gamma = \delta = 1$), if exactly one element $a_{ij} > 0$ (for $i < m$ is $\gamma = i$, for $i = m$ is $\gamma = i-1$, for $j < n$ is $\delta = j$, for $j = n$ is $\delta = j-1$), if exactly two elements $a_{ij} > 0$ and $a_{ij+1} > 0$ or $a_{ij} > 0$ and $a_{i+1j} > 0$ (for $i < m$ is $\gamma = i$, for $i = m$ is $\gamma = i-1$, for $j < n$ is $\delta = j$, for $j = n$ is $\delta = j-1$). Second, it is of type $\gamma\delta$-row, if exactly two elements $a_{ij} > 0$ and $a_{ik} > 0$ such that $i \leq m$, $j < n$ and $j+1 < k \leq n$ (for $i < m$ is $\gamma = i$, for $i = m$ is $\gamma = i-1$, $\delta = j$). Finally, it is of type $\gamma\delta$-column, if exactly two elements $a_{ij} > 0$ and $a_{kj} > 0$ such that $i < m$, $j \leq n$ and $i+1 < k \leq m$ ($\gamma = i$, for $j < n$ is $\delta = j$, for $j = n$ is $\delta = j-1$).
A $\gamma\delta$ matrix is a special case of a $\gamma\delta$-row matrix and $\gamma\delta$-column matrix with the same γ and δ.

Remark 2. For arbitrary $h \in H = \{a_{ij};\ i,j \in N\}$, $h > 0$, the threshold digraph $G(A,h)$ of a matrix A of type $\gamma\delta$ or $\gamma\delta$-row or $\gamma\delta$-column contains only arcs crossing node γ or δ. Consequently it can contain at most three cycles, namely, loops on nodes γ and $\gamma+1$ (for a $\gamma\delta$-row matrix) or δ and $\delta+1$ (for a $\gamma\delta$-column matrix) and a cycle of length two $c = (\gamma,\ \gamma+1,\ \gamma)$ (for a $\gamma\delta$-row matrix) or $c = (\delta,\ \delta+1,\ \delta)$ (for a $\gamma\delta$-column matrix). Moreover, $G(A,h)$ has at most two non-trivial strongly connected components only.

Algorithm **Type of matrix**

Input. $A \in B(n,n)$.
Output. '$\gamma\delta$' in variable $type$, if A is a $\gamma\delta$ matrix with corresponding values in variables γ and δ, '$\gamma\delta$-row' in variable $type$, if A is a $\gamma\delta$-row matrix with corresponding values in variables γ and δ, '$\gamma\delta$-column' in variable $type$, if A is a $\gamma\delta$-column matrix with corresponding values in variables γ and δ, 'null' in variable $type$, if A is no one of the considered types.

begin

1. $gamma = n$; $delta := n$; $p := 1$; $q := 1$; $isZero$:='true'; $i := 1$; $j := 1$;
2. **If** $a_{ij} = 0$ **then go to** step 8;
3. $isZero$:='false';
4. **If** $j < \delta$ **then** $delta := j$;
5. **If** $i < gamma$ **then** $gamma := i$;
6. **If** $p < i$ **then** $p := i$;
7. **If** $q < j$ **then** $q := j$;
8. $j := j + 1$;
9. **If** $j = n+1$ **then** $j := 1$; $i := i + 1$;
10. **If** $i < n+1$ **then go to** step 2;
11. **If** $isZero$ ='true' **then** $delta := 1$; $gamma := 1$; $type$:='$\gamma\delta$'; **go to** end;
12. **If** $delta = n$ **then** $delta := delta - 1$;
13. **If** $gamma = n$ **then** $gamma := gamma - 1$;
14. $colSize := q - delta + 1$; $rowSize := p - gamma + 1$;
15. **If** $colSpan > 2$ **and** $rowSpan > 2$ **then** $type$:='null'; **go to** end;
16. **If** $colSpan > 2$ **and** $rowSpan \leq 2$ **then** $type$:='$\gamma\delta$-row'; **go to** end;
17. **If** $colSpan \leq 2$ **and** $rowSpan > 2$ **then** $type$:='$\gamma\delta$-column'; **go to** end;
18. **If** $colSpan \leq 2$ **and** $rowSpan \leq 2$ **then** $type$:='$\gamma\delta$'.

end

Theorem 2. *Let $A \in B(n,n)$. The Algorithm* **Type of matrix** *correctly checks in $O(n^2)$ time whether the matrix A is a matrix of special type $\gamma\delta$ or $\gamma\delta$-row or $\gamma\delta$-column.*

Proof. To find the smallest value γ such that $a_{\gamma j} > 0$, the smallest value δ such that $a_{i\delta} > 0$, the biggest number p such that $a_{pj} > 0$ and the biggest number q such that $a_{iq} > 0$ takes $O(n^2)$. To compute the values of $rowSpan$ and $colSpan$, which represents the distance of elements with values bigger than zero in row γ or column δ, respectively, together with estimation of the type of the given matrix takes $O(1)$ time. □

Theorem 3. *There is an algorithm with computational complexity $O(n^2)$ for verifying the Monge property of a given matrix $A \in B(n,n)$ of special type $\gamma\delta$ or $\gamma\delta$-row or $\gamma\delta$-column.*

Proof. Using Definition 2 it is necessary to check only two rows (γ and $\gamma+1$) or two columns (δ and $\delta+1$), respectively, of the given matrix. Thus the computational complexity is $O(n^2)$. □

Lemma 2. [6] Let $A \in B(n,n)$ be a Monge matrix. Let $h \in H$. Let for $i,\ k \in N$ be the loops (i,i) and (k,k) in the digraph $G(A,h)$. Then the nodes i and k are in the same non-trivial strongly connected component $\mathcal{K}$ of $G(A,h)$.

Theorem 4. *Let $A \in B(n,n)$ be a Monge matrix of type $\gamma\delta$-row or $\gamma\delta$-column, respectively. Then A is robust if and only if for each $h \in H$ holds:*

$$a_{\gamma\gamma+1} \otimes a_{\gamma+1\gamma} \geq h \Longrightarrow a_{\gamma\gamma} \oplus a_{\gamma+1\gamma+1} \geq h \quad \textit{or}$$

$$a_{\delta\delta+1} \otimes a_{\delta+1\delta} \geq h \Longrightarrow a_{\delta\delta} \oplus a_{\delta+1\delta+1} \geq h,$$

respectively.

Proof. Let A be a Monge matrix of type $\gamma\delta$-row. Let us assume that A is robust. Let $h \in H$ be arbitrary but fixed. Let $a_{\gamma\gamma+1} \otimes a_{\gamma+1\gamma} \geq h$ and let $a_{\gamma\gamma} \oplus a_{\gamma+1\gamma+1} < h$. Thus by Remark 2 there is only one non-trivial strongly connected component of the digraph $G(A, h)$ and this contains a cycle of length two and no loop. This is by Theorem 1 a contradiction with robustness of the considered matrix A.
The converse implication is a consequence of Remark 2, Lemma 2 and Theorem 1. □

Theorem 5. *There is an algorithm with computational complexity $O(n^2)$ for verifying the type and the Monge property of a given matrix $A \in B(n, n)$ and checking the robustness.*

Proof. Due to Theorem 2 the verifying the type of the given matrix takes $O(n^2)$ time. Using Theorem 3 the verifying of the Monge property takes another $O(n^2)$ time. Since H contains at most $2n$ different elements and to verify the inequalities takes $O(1)$ time for each $h \in H$ the entire computational complexity is $O(n^2)$. □

4 Robustness of $\gamma\delta$-row and $\gamma\delta$-column Monge matrix with inexact data

Robustness of interval fuzzy matrices of type $\gamma\delta$-row and $\gamma\delta$-column with the Monge property will be studied in this section. 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 between a lower bound matrix $\underline{A}$ and an upper bound matrix $\overline{A}$. Hence, the question is 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 $\gamma\delta$ or $\gamma\delta$-row or $\gamma\delta$-column, respectively, if $\overline{A}$ is of type $\gamma\delta$ or $\gamma\delta$-row or $\gamma\delta$-column, respectively, with same $\gamma,\ \delta \in N$.

Definition 8. An interval Monge matrix $\mathbf{A}^M$ is called an interval Monge matrix of type $\gamma\delta$ or $\gamma\delta$-row or $\gamma\delta$-column, respectively, if $\mathbf{A} = [\underline{A}, \overline{A}]$ is of type $\gamma\delta$ or $\gamma\delta$-row or $\gamma\delta$-column, respectively.

Remark 3. Since $\underline{A} \leq \overline{A}$ is the matrix $\overline{A}$ significant for determination of the type of the matrix $\mathbf{A}$. Consequently are all matrices from $\mathbf{A}$ considered to be of the same type as $\overline{A}$ with same γ and δ, although they can have different parameters, in fact.

Since the computational complexity of the algorithm for checking the equivalent conditions for a fuzzy interval matrix to be possibly robust described in [9] is $O(n^5)$ and for checking the universal robustness there is even no polynomial algorithm, it is proper to investigate special classes of matrices to find more effective algorithms. Hence, we will prove sufficient condition for possible robustness of an interval matrix of type $\gamma\delta$-row or $\gamma\delta$-column, respectively, with computational complexity $O(n^2)$. Moreover, we will prove equivalent conditions for an interval matrix of type $\gamma\delta$-row or $\gamma\delta$-column, respectively, 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 6. *An interval Monge matrix $\mathbf{A}^M$ of type $\gamma\delta$-row is possibly robust if $\overline{A}$ is robust.*

Proof. The assertion follows by Definition 5. □

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 possible robustness of $\mathbf{A}^M$.*

Proof. First, we shall verify the type of the interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ by verifying the type of matrices $\underline{A}$, $\overline{A}$ using Algorithm **Type of the matrix**. Due to Theorem 2 it takes $O(n^2)$ time. Second, we shall check the Monge property of the lower bound matrix $\underline{A}$ and the upper bound matrix $\overline{A}$. Due to Theorem 3 it takes $O(n^2)$ time. Finally, to verify the robustness of $\overline{A}$ by Theorem 5 takes $O(n^2)$ time, hence the entire computational complexity is $O(n^2)$. □

Universal robustness

Theorem 8. *An interval Monge matrix $\mathbf{A}^M$ of type $\gamma\delta$-row with $\gamma = \delta$ is universally robust if and only if $\underline{a}_{\gamma\gamma} \oplus \underline{a}_{\gamma+1\gamma+1} \geq \overline{a}_{\gamma\gamma+1} \otimes \overline{a}_{\gamma+1\gamma}$.*

Proof. First, we prove the sufficient condition. Let $A \in \mathbf{A}^M$. Let $h > 0$. There are two possibilities. First, the trivial case, the threshold digraph $G(A, h)$ contains no cycle of positive length. Thus $G(A, h)$ contains no non-trivial strongly connected component. Second, $G(A, h)$ contains at lest one cycle of positive length. Due to assumption it contains loop on node γ or/and node $\gamma + 1$. Since A is a Monge matrix following Lemma 2 loops lie in the same strongly connected component. Due to Remark 2 and Theorem 1 A is robust, hence $\mathbf{A}^M$ is universally robust.
Now, we prove the necessary condition. Let $\mathbf{A}^M$ be universally robust. We shall prove that $\underline{a}_{\gamma\gamma} \oplus \underline{a}_{\gamma+1\gamma+1} \geq \overline{a}_{\gamma\gamma+1} \otimes \overline{a}_{\gamma+1\gamma}$. Let $\underline{a}_{\gamma\gamma} \oplus \underline{a}_{\gamma+1\gamma+1} < \overline{a}_{\gamma\gamma+1} \otimes \overline{a}_{\gamma+1\gamma}$. Then there is a matrix $A^\star \in \mathbf{A}^M$ defined as follows

$$a^\star_{ij} = \begin{cases} \underline{a}_{\gamma\gamma} & \text{if } i = j = \gamma, \\ \underline{a}_{\gamma+1\gamma+1} & \text{if } i = j = \gamma + 1, \\ \overline{a}_{ij} & \text{otherwise.} \end{cases} \tag{4}$$

First, we shall prove that $A^\star$ is a Monge matrix. Thus we shall prove that $a^\star_{\gamma\gamma} \otimes a^\star_{\gamma+1l} \leq a^\star_{\gamma l} \otimes a^\star_{\gamma+1\gamma}$ for $l \geq \gamma+1$. Let us consider $l = \gamma+1$. Using definition of the matrix $A^\star$ the inequality turns to $\underline{a}_{\gamma\gamma} \otimes \underline{a}_{\gamma+1\gamma+1} \leq \overline{a}_{\gamma\gamma+1} \otimes \overline{a}_{\gamma+1\gamma}$ and follows by the assumption. For $l > \gamma + 1$ the inequality turns to $\underline{a}_{\gamma\gamma} \otimes \overline{a}_{\gamma+1l} \leq \overline{a}_{\gamma l} \otimes \overline{a}_{\gamma+1\gamma}$ and follows by $\underline{a}_{\gamma\gamma} \leq \overline{a}_{\gamma\gamma}$ and the Monge property of the matrix $\overline{A}$. Further we shall prove that $a^\star_{\gamma\gamma+1} \otimes a^\star_{\gamma+1l} \leq a^\star_{\gamma l} \otimes a^\star_{\gamma+1\gamma+1}$ for $l > \gamma + 1$. The inequality turns to $\overline{a}_{\gamma\gamma+1} \otimes \overline{a}_{\gamma+1l} \leq \overline{a}_{\gamma l} \otimes \overline{a}_{\gamma+1\gamma+1}$ and follows by the Monge property of the matrix $\overline{A}$.
Second, we shall prove that $A^\star$ is not robust. Let us consider the threshold digraph $G(A^\star, h)$ for $h = \overline{a}_{\gamma\gamma+1} \otimes \overline{a}_{\gamma+1\gamma}$. The digraph contains a non-trivial strongly connected component with a cycle of length two connecting nodes γ and $\gamma + 1$. Since $a^\star_{\gamma\gamma} \oplus a^\star_{\gamma+1\gamma+1} = \underline{a}_{\gamma\gamma} \oplus \underline{a}_{\gamma+1\gamma+1} < \overline{a}_{\gamma\gamma+1} \otimes \overline{a}_{\gamma+1\gamma} = a^\star_{\gamma\gamma+1} \otimes a^\star_{\gamma+1\gamma}$ the component is without any loop. Thus by Theorem 1 the matrix $A^\star$ is not robust. What is a contradiction with universal robustness of $\mathbf{A}^M$. □

Due to Remark 2 we can formulate and prove by analogy the necessary and sufficient condition for a $\gamma\delta$-column interval Monge matrix to be universally robust.

Theorem 9. *An interval Monge matrix $\mathbf{A}^M$ of type $\gamma\delta$-column is universally robust if and only if $\underline{a}_{\delta\delta} \oplus \underline{a}_{\delta+1\delta+1} \geq \overline{a}_{\delta\delta+1} \otimes \overline{a}_{\delta+1\delta}$.*

Theorem 10. *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 of $\mathbf{A}^M$.*

Proof. First, we shall verify the type of the interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ by verifying the type of matrices $\underline{A}$, $\overline{A}$ using Algorithm **Type of the matrix**. Due to Theorem 2 it takes $O(n^2)$ time. Second, we shall check the Monge property of the lower bound matrix $\underline{A}$ and the upper bound matrix $\overline{A}$. Due to Theorem 3 it takes $O(n^2)$ time. Finally, to verify the inequality takes $O(1)$ 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 $\gamma\delta$-row using Theorem 8.

Example 1. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A}$, $\overline{A} \in B(5, 5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 2 & 3 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \overline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 4 & 2 & 3 \\ 0 & 4 & 4 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \qquad A^\star = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 4 & 2 & 3 \\ 0 & 4 & 3 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrix $\overline{A}$ is of type $\gamma\delta$-row with $\gamma = 2$ and $\delta = 2$. Moreover, both $\underline{A}$ and $\overline{A}$ are Monge matrices. Hence $\mathbf{A}^M$ is a interval Monge matrix of type $\gamma\delta$-row with $\gamma = 2$ and $\delta = 2$. Since $\underline{a}_{22} \oplus \underline{a}_{33} = 3 < \overline{a}_{23} \otimes \overline{a}_{32} = 4$ the matrix $\mathbf{A}^M$ is not universally robust. In fact, there is a Monge matrix $A^\star$ defined by (4) which is not robust, namely the threshold digraph $G(A^\star, 4)$ 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.

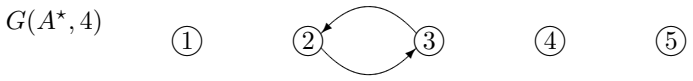


Figure 1 Universal robustness in non-robust case

Example 2. Let us consider an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ with bounds $\underline{A}$, $\overline{A} \in B(5, 5)$

$$\underline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 2 & 4 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \ \overline{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 4 & 2 & 3 \\ 0 & 4 & 4 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

A slight modification of the previous example on position $\underline{a}_{33}$ gives a positive answer. Since $\underline{a}_{22} \oplus \underline{a}_{33} = 4 \geq \overline{a}_{23} \otimes \overline{a}_{32} = 4$ the interval Monge matrix $\mathbf{A}^M$ of type $\gamma\delta$-row is universally robust.

5 Conclusion

The aim of this paper is to find equivalent conditions for robustness of Monge fuzzy matrices of type $\gamma\delta$, $\gamma\delta$-row and $\gamma\delta$-column with fixed data and sufficient or equivalent conditions for possible robustness and universal robustness of interval Monge fuzzy matrices of type $\gamma\delta$, $\gamma\delta$-row and $\gamma\delta$-column, respectively.

The robustness of the matrix A is related to the solution of the eigenproblem $A \otimes x(r) = x(r)$, where an eigenvector $x(r)$ represents the steady state and can be computed at time $r + 1$ by recursion $x(r + 1) = A \otimes x(r)$. This can reflect the following economic application for instance. In a company several projects should be evaluated. 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}.

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 Monge matrices in fuzzy algebra, In: *Proceedings of* 32^{nd} *Int. Conference Mathematical Methods in Economics 2014*, Olomouc, 2014, 679–684.

[7] 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.

[8] 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.

[9] Molnárová, M., Myšková, H. and Plavka, J.: The robustness of interval fuzzy matrices, *DAM*, Volume **438** (2013), 3350–3364.

[10] Plavka, J.: On the $O(n^3)$ algorithm for checking the strong robustness of interval fuzzy matrices, *DAM*, Volume **160** (2012), 640–647.

[11] Plavka, J. and Sergeev, S.: Characterizing matrices with X-simple image eigenspace in max-min algebra, *Kybernetika*, Volume **52** Issue 4 (2016), 497–513.

[12] Plavka, J. and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *DAM*, Volume **159** Issue 5 (2011), 381–388.

Analysis of the performance of control charts in the presence of assignable causes

Pablo José Moya Fernández[1], Juan Francisco Muñoz Rosas[2], Encarnación Álvarez Verdejo [3], Hana Skalská [4]

Abstract. The main objective of control charts is to evaluate a production process. The process is stable or in-control when the variability of the production process is only produced by common causes. However, if the variability is produced by assignable causes it is said that the process is in a situation of out-of-control. An advantage of control charts is the detection and identification of assignable causes within the production process. It is important to know the situation of a process because alternative quantitative techniques in the field of statistical quality control are based upon a process under control. Simulation studies are carried out to analyze the performance of control charts under different scenarios. In particular, we first consider that the processes have a correct operation. Second, we consider processes that operates with quality characteristics that do not have the required assumption, and this issue has an impact on the proportion of produced articles. Third, we consider processes that suffer from a change on the performance of the production volume.

Keywords: Simulation study, control limits, variance, statistical process control.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Control charts are one of the main techniques of statistical process control and have been used by many companies since they are especially useful thanks to the benefits they provide [2, 6]. The main objective of control charts is to control the parameters associated with the quality characteristic. In general, companies are interested in controlling the variability of the process. In line with this issue, the control charts monitor the production process to detect causes of assignable variation, which allows to apply corrective actions as soon as possible, thus minimizing the losses generated by the manufacture of non-conforming items. Therefore, the control charts allow to identify the status of the process.

The control charts are a graphic representation that includes a chronological series of the statistic corresponding to the process parameter to be controlled, a central line and the control limits. The chronological series of the statistic is determined from random samples of products that have been selected in a step prior to the elaboration of the control chart. The value of the statistic for each sample will be represented by a point, which will be joined with the points of the following samples by straight lines between each two consecutive points. The central line (CL) of the control chart is a horizontal line that corresponds to the expected value of the estimator of the population parameter under study. The upper limit of control (ULC) and the lower control limit (LCL) are two horizontal lines at a predetermined distance from the central line that delimit the values that the statistician can present when the process is under control [1, 5, 9, 10, 13].

The process is in-control if all the points represented in the control chart are located in the area delimited by the control limits and are distributed randomly, without ascending or descending streaks of points. When the process is in this situation, the process is stable and predictable, that is, the variability within it is due exclusively to fortuitous causes and the uniformity between articles is maximum. Besterfield [1] lists some of the benefits that arise when the process is in this state. For example, the possibility of minimizing the cost per inspection, since fewer samples are needed to evaluate the quality. Conversely, if any statistic value is above the UCL or below the LCL, the process will be in an out-of-control state. In addition, although all the statistic's values are located in the area delimited by both limits, if these are not distributed randomly, the process is also out-of-control. Some useful rules can be consulted to determine if the process is out-of-control in [1, 7, 10]. Some very common rules are the presence of ascending or descending streaks of points, four of five consecutive points located at a distance greater

[1]Universidad de Granada, Dpto. Métodos Cuantitativos para la Economía y la Empresa, Spain, pjmoyafernandez@ugr.es
[2]Universidad de Granada, Dpto. Métodos Cuantitativos para la Economía y la Empresa, Spain, jfmunoz@ugr.es
[3]Universidad de Granada, Dpto. Métodos Cuantitativos para la Economía y la Empresa, Spain, encarniav@ugr.es
[4]University of Hradec Králové, Department of Informatics and Quantitative Methods, Czech Republic, hana.skalska@uhk.cz

than the standard deviation of the quality characteristic of the CL, a series of eight consecutive points on one side of the CL, fourteen points alternating up and down or an unusual pattern in the data.

The main objective of this work is to analyze the performance of the control charts by means of a simulation study. We consider different scenarios which correspond to situations may arise in practice:

a) We suppose a company that operates with a process in which a certain machine is composed of a series of gears. A particular gear contains a tooth of unsuitable dimensions, resulting from excessive wear. Some articles are influenced by the defective tooth of this machine. We assume that this fact directly influences the quality characteristic of the product, resulting in a higher value than is usual. In the control chart we expect the higher value of statistic corresponding to samples that include articles produced with the intervention of the defective tooth. This point could be located above UCL or will represent a pattern of behaviour in the control chart. Both cases can indicate that the process is not under control and can be qualified as a state out of control. That is, a part of the variability of the process is produced by an assignable cause: the tooth of unsuitable dimensions.

b) A company working with a process involves a machine that, after producing a number of items, overheats. While this machine is overheated, the process produces articles with a certain alteration in their quality characteristic, such as a slightly longer length. The control chart corresponding to this process can indicate a pattern or a streak on more or less long series of points located at a certain distance from the central line of the control chart in the chronological series of the sample statistic. Points outside the range the control limits and both above mentioned scenarios are considered as out-of-control and the assignable cause is the defective machine.

This paper is organized as follows. In section 2, we describe the simulation study carried out to analyze the performance of control charts in the presence of assignable causes. Several processes with different characteristics have been considered. Note that these processes may have assignable causes with different effects on the quality characteristic and which may arise in practice. This simulation study has been carried out with the R software [12]. In section 3, the most relevant results are presented and discussed. The main conclusions derived from this simulation study are summarized in section 4.

2 Methodology

The analysis of the performance of the control charts is carried out using a simulation study, which consists in drawing the $\bar{x}$ and s control charts for different processes. In each process, production is stored in lots, from which a sample of items is extracted. In each item, the quality characteristic is measured and the statistic of the sample is calculated. Afterwards, the control chart is elaborated. We assume that the process associated parameters are known. In the case that the unknown parameters, they should be estimated using the information contained in product samples. It is very important to select the samples when it is certain that the process is in an in-control state [3, 8, 15]. In order to illustrate this fact, the control limits estimated using the samples selected for the control chart have been included. In this way, the impact on the control charts of estimating the parameters when the process is out-of-control is analyzed. Some references where the determination of control limits with known and unknown parameters can be consulted [4, 11, 14].

Let $x \to N(\mu, \sigma)$ be a quality characteristic associated to a process, where μ is the true process mean and σ is true process standard deviation. We assume m random samples with size n, where x_{ij} denote the observed value of the quality characteristic for the jth product, with $j = 1, ..., n$, in the ith sample, with $i = 1, ..., m$. The process can be monitored by plotting the sample means and sample standard deviations

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}, \quad s_i = \left(\frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2\right)^{1/2}, \qquad i = 1, ..., m. \tag{1}$$

Consider the case of known parameters, the control limits for $\bar{x}$ control chart are defined as

$$LCL_\mu = \mu - 3\frac{\sigma}{\sqrt{n}}; \quad CL_\mu = \mu; \quad UCL_\mu = \mu + 3\frac{\sigma}{\sqrt{n}} \tag{2}$$

and for s control chart as

$$LCL_\sigma = c_4\sigma - 3\sigma\sqrt{1 - c_4^2}; \quad CL_\sigma = c_4\sigma; \quad UCL_\sigma = c_4\sigma + 3\sigma\sqrt{1 - c_4^2}, \tag{3}$$

where c_4 is a function that depends on the sample size [4].

If we consider the case of unknown parameters, the control limits for $\bar{x}$ and s control chart are defined as

$$LCL_{\mu_u} = \bar{\bar{\mu}} - 3\frac{\bar{s}}{\sqrt{n}}; \quad CL_\mu = \bar{\bar{\mu}}; \quad UCL_{\mu_u} = \bar{\bar{\mu}} + 3\frac{\bar{s}}{\sqrt{n}} \tag{4}$$

and

$$LCL_{\sigma_u} = \bar{s} - 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2}; \quad CL_{\sigma_u} = \bar{s}; \quad UCL_{\sigma_u} = \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} \tag{5}$$

respectively, where $\bar{\bar{\mu}} = m^{-1}\sum_{i=1}^{m}\bar{x}_i$ is the grand sample mean and $\bar{s} = m^{-1}\sum_{i=1}^{m} s_i$ is the average of the m standard deviations.

Three different processes have been considered in this simulation study. The first scenario corresponds to a productive process that operates correctly. In the second scenario, the production process operates with a certain alteration in the quality characteristic, and which only affects a proportion p^* of the articles produced. In this second scenario, we will assume that each lot contains that proportion p^* of items. Finally, the third scenario corresponds to a productive process that, from a certain volume of production, changes its performance. Since lots storage is chronological, we assume that only a certain number of lots contain items produced by the process when the incidence is present. In addition, we will assume that the proportion of lots with item with this type of incidences is also given p^*. For example, if we assume $p^* = 0.2$ and lots of size 100, in the second scenario each lot stores 20 items with incidence. Under the third scenario and in a cyclical way, 20 % of the lots will contain item with incidents, i.e., the production will be stored in four consecutive lots with articles without incidence, while the next batch will contain articles with incidence. Note that the case studies presented in the previous section can be identified with the second and third scenarios, respectively.

Therefore, two different types of production are presented. On the one hand, the production corresponding to a process when it operates without any incidence and, on the other hand, the production result of a process when it is affected by some incidence. We assume that the variable that represents the quality characteristic of the articles produced by a process without incidences has a normal distribution with mean μ and standard deviation σ. Consequently, the quality characteristic corresponding to production with incidences follows a normal distribution with mean $\mu + \epsilon_\mu$ and standard deviation $\sigma + \epsilon_\sigma$. In the simulation carried out in this work, control charts have been drawn up in which 50 samples of size 5 are represented and the control limits are defined using the USA method [10]. To simulate the 3 previously defined scenarios, lots of size 100 and an incidences proportion of 20% ($p^* = 0.2$) for the second and third scenarios have been assumed. We consider the values $\mu = 10$ and $\sigma = 1$ for the parameters associated with the production without incidents. In addition, in order to represent the changes in production discussed, the values $\epsilon_\mu = \{0, 1, 2, 3\}$ and $\epsilon_\sigma = \{0, 0.5, 1, 2\}$ and all its possible combinations are considered. Note that the objective of this work is to evaluate the impact of assignable causes in the process with different effects, both in intensity and in form.

3 Results derived from simulation and discussion

The most relevant results derived from the simulation study are shown in the Figures1-3. Each Figure contains the control charts corresponding to the three scenarios. Note that the first scenario is represented in all the figures as a reference, since it is characterized as a process under control that does not suffer variation.

Figure 1 contains the $\bar{x}$ and s control charts when the incidence in the process only affects the standard deviation of the product quality characteristic. In this case, we can see in the s control chart sufficient indications to think that, in the second as in the third scenario, the process is out-of-control. In the second scenario, the sample standard deviation of the samples number 17, 23 and 24 are outside the range delimited by the control limits. Something similar is observed in the sample number 15 corresponding to the third scenario. On the other hand, the $\bar{x}$ control charts show for both scenarios several points close to the control limits, which may be an indication that the process is out-of-control.

Figure 2 contains the $\bar{x}$ and s control charts when the incidence in the process has influence only on the average of the quality characteristic, i.e., $\epsilon_\mu = 2$ and $\epsilon_\sigma = 0$. We observed that, in the third scenario, the total of the sample means corresponding to the production with incidence in the process are above UCL, while the standard deviation seems to be under control. The process of the second scenario is also in an out-of-control situation. The standard deviation of sample number 29 is outside the range defined by the control limits. In addition, two points corresponding to samples 23 and 44 are outside the control limits of the control chart for the mean and we observed in samples 47 and 50 another situation to identify the process as out-of-control, one or more points to a warning or control limit.

The $\bar{x}$ and s control charts assuming values $\epsilon_\mu = 2$ and $\epsilon_\sigma = 0.5$ are shown in Figure 3. In this case the incidence in the process affects both the mean and the standard deviation of the product quality characteristic. The control charts contain indications that the second and third scenario processes are out-of-control. Among these, highlights the existence of points outside the control limits of the $\bar{x}$ and s control chart in both processes.

Additionally, the estimated control limits have been represented in each control chart. The information contained in the samples represented in the control chart has been used to carry out the estimation of the control limits. Therefore, the control limits in the second and third scenarios have been estimated when the process is out-of-

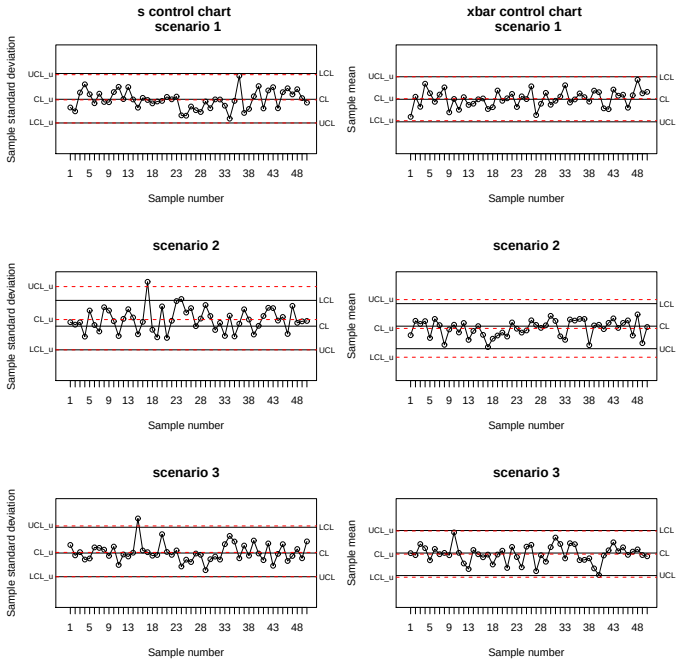


Figure 1 $\bar{x}$ and s control charts of a process. Samples are selected from a Normal distribution $\mu = 10$ and $\sigma = 1$. The production with incidence represents a proportion of 0.2 and its quality characteristic is affected by the values $\epsilon_\mu = 0$ and $\epsilon_\sigma = 0.5$.

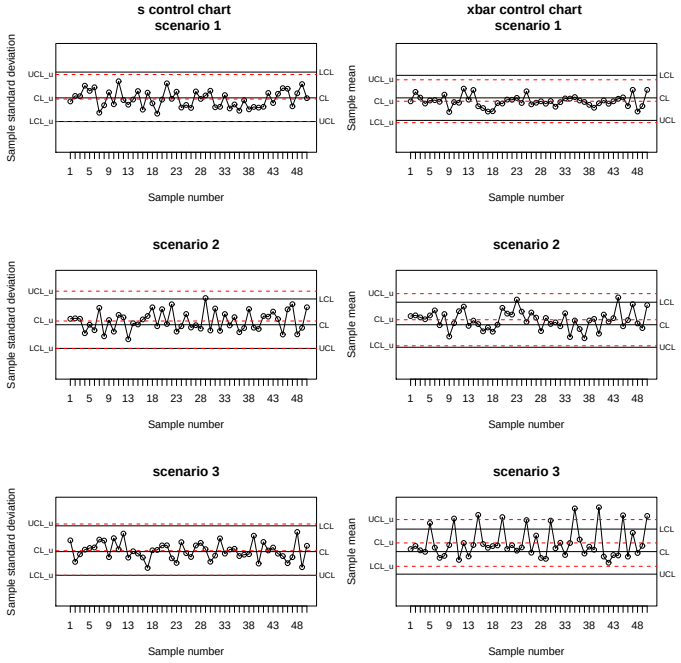


Figure 2 $\bar{x}$ and s control charts of a process. Samples are selected from a Normal distribution $\mu = 10$ and $\sigma = 1$. The production with incidence represents a proportion of 0.2 and its quality characteristic is affected by the values $\epsilon_\mu = 2$ and $\epsilon_\sigma = 0$.

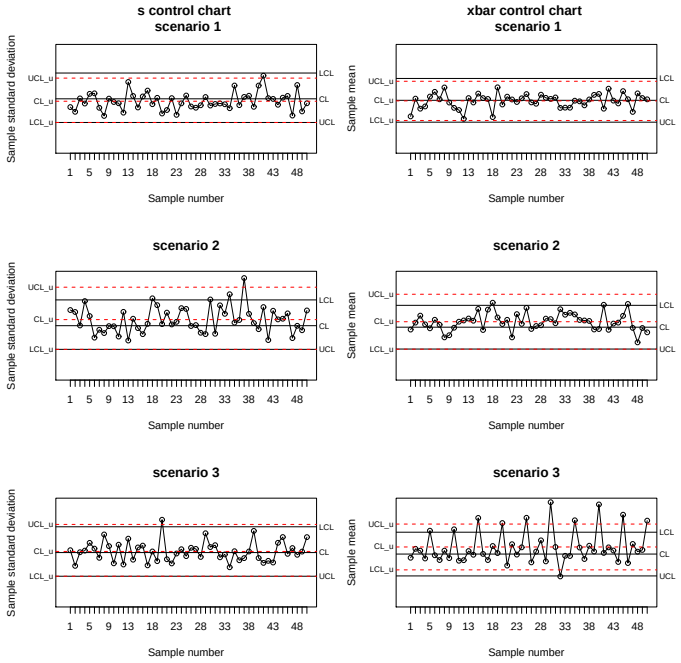


Figure 3 $\bar{x}$ and s control charts of a process. Samples are selected from a Normal distribution $\mu = 10$ and $\sigma = 1$. The production with incidence represents a proportion of 0.2 and its quality characteristic is affected by the values $\epsilon_\mu = 2$ and $\epsilon_\sigma = 0.5$.

control situation. We observed that, in general, the estimated control limits are closer to the control limits when the parameters are known when the process is in a control situation. In this situation, the conclusions obtained from the control charts are similar. Otherwise, the estimated control limits may be located far from the control limits and the conclusions obtained in the control charts may be different. For example, an assignable cause is present in the process of the second scenario that affects a proportion of 0.2. We observe in Figure 2 the control charts when this assignable cause affects only the average of the process. Considering that the parameters are known and therefore the control limits, we observe several points outside the control limits in both graphs. However, if we consider the unknown parameters and the estimated control limits, all points of both graphs are within the control limits.

4 Conclusion

In this work we have analyzed the performance of control charts in the presence of assignable cases. To achieve this objective, a simulation study was carried out in which the control charts corresponding to different processes were prepared. We have considered different processes that presented variability due to assignable causes, which affected production with different consequences, frequency and intensity. In each scenario, the $\bar{x}$ and s control charts have been developed assuming known parameters. Additionally, the control limits have been represented assuming unknown parameters with the objective of analyzing the impact that their estimate has when the process is in an out-of-control situation.

According to the obtained results, we may affirm that the control charts are an highly effective tool in the detection of assignable causes in the process, since they have been able to identify indications that the process is out-of-control independently of the characteristics of the consequences produced for the assignable cause. This is really useful, since it allows to apply corrective actions as soon as possible, thus minimizing the losses generated by the manufacture of non-conforming articles. In addition, control charts report numerous advantages. For example, improve the productivity of the company, reduce the costs derived from an over-intervention in the process or provide information on the capacity of the process.

On the other hand, the importance of estimating the parameters, and with them the limits of control, when the situation of the process is in-control has been demonstrated. The estimation of the parameters associated with the process using samples from a process out-of-control could considerably alter the position of the control limits. Whereas if the estimation of the parameters is done using samples from a process in-control, the estimated control limits will be close to the control limits. In conclusion, if the control limits are estimated using samples from out-of-control processes, the control charts may lose effectiveness in the detection of variability due to assignable causes in the process, which could lead to erroneous conclusions or confusion when drawing conclusions.

References

[1] Besterfield, D. (2009). *Control de Calidad.(9th edn)*. Prentice Hall. Hispanoamericana S.A.

[2] Besterfield, D. (2014). *Quality improvement.(9th edn)*. Harlow: Pearson Education.

[3] Chakraborti, S.,Human, S. W. & Grahan, M. A. (2008). Phase I statistical process control charts: an overview and some results. *Quality Engineering*, 21(1), 52–62.

[4] Chen, G. (1997). The mean and standard deviation of the run length distribution of X charts when control limits are estimated. *Statistica Sinica*, 7, 789–798.

[5] Dovich, R. A. (1992). *Quality engineering statistics*. ASQ Quality Press.

[6] Evans, J. R. & Lindsay, W. M. (1999). *The management and control of quality.(8th edn)*. South Western College.

[7] Hurwitz, S. & Mathur, M. (1992). A very simple set of process control rules. *Quality Engineering*, 5(1), 21–29.

[8] Jensen, W. A., Jones-Farmer, L. A., Champ, C. W. & Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology*, 38(4), 349–364.

[9] Mitra, A. (2008). *Fundamentals of quality control and improvement*. John Wiley & Sons.

[10] Montgomery, D. C. (2009). *Statistical Quality Control. A modern introducction.(6th edn)*. New York, Wiley.

[11] Muñoz-Rosas, J. F., Álvarez-Verdejo, E., Pérez-Aróstegui, M. N. & Gutiérrez-Gutiérrez, L. (2016). Empirical Comparisons of Xbar Charts when Control Limits are Estimated. *Quality and Reliability Engineering International*, 32(2), 453–464.

[12] Team, R. C. (2017). *Team, R. C. (2017). R: A language and environment for statistical computing.*. R Foundation for Statistical Computing, Vienna, Austria. 2016.

[13] Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. ASQ Quality Press.

[14] Vardeman, S. B. (1999). A brief tutorial on the estimation of the process standard deviation. *IIE transactions*, 31(6), 503–507.

[15] Wheeler, D. J. (1995). *Advanced topics in statistical process control*. Knoxville, TN: SPC press.

Strongly Tolerance Solvability of Max-plus Matrix Equations

Helena Myšková[1]

Abstract. Max-plus algebra is an algebraic structure, in which classical addition and multiplication are replaced by maximum and addition, respectively. Behavior of discrete event systems, in which the individual components move from event to event rather than varying continuously through time, is often described by systems of max-plus linear equations or matrix equations. Discrete dynamic systems can be studied using max-plus matrix operations. It often happens that a max-plus matrix equation with exact data is unsolvable. Therefore, we replace matrix elements with intervals of possible values. In this way, we obtain an interval max-plus matrix equation. Several types of solvability of interval max-plus matrix equations have been studied yet. In this paper, we prove the necessary and sufficient condition for an interval max-plus matrix equation to be strongly tolerance solvable and provide an algorithm for checking whether the given interval matrix equation is strongly tolerance solvable.

Keywords: max-plus algebra, interval matrix, matrix equation, strongly tolerance solvability

JEL classification: C02
AMS classification: 15A18; 15A80; 65G30

1 Motivation

Behaviour of discrete event systems, in which the individual components move from event to event rather than varying continuously through time, is often described by systems of linear equations or by matrix equations. Discrete dynamic systems and related algebraic structures were studied using max-plus matrix operations in [2, 3]. In the last decades, significant effort has been developed to study systems of max-plus linear equations in the form $A \otimes x = b$, where A is a matrix, b and x are vectors of compatible dimensions. Systems of linear equations over max-plus algebra are used in several branches of applied mathematics. Among interesting real-life applications let us mention e.g. a large scale model of Dutch railway network or synchronizing traffic lights in Delfts [12]. In the last two decades, interval systems of the form $\boldsymbol{A} \otimes x = \boldsymbol{b}$ have been studied, for details see [2, 4, 5].

In this paper, we shall deal with interval max-plus matrix equations of the form $\boldsymbol{A} \otimes X \otimes \boldsymbol{C} = \boldsymbol{B}$, where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are given interval matrices of suitable sizes a X is an unknown matrix. Several solvability concepts have been studied in [10], [11]. The following example is the shortened version of an example given in [10].

Example 1. Let us consider a situation, in which passengers from places $P_1, P_2, \ldots, P_m$ want to transfer to holiday destinations $D_1, D_2, \ldots, D_r$. Different transportation means provide transporting passengers from places $P_1, P_2, \ldots, P_m$ to airport terminals $T_1, T_2, \ldots, T_s$. We assume that the connection between P_i and T_l is possible only via one of the check points $Q_1, Q_2, \ldots, Q_n$.

Denote by a_{ij} (c_{lk}) the times needed for transportation and carrying out the formalities on the connection from P_i to Q_j (from T_l to D_k). If there is no connection from P_i to Q_j (from T_l to D_k) we put $a_{ij} = -\infty$ ($c_{lk} = -\infty$). If the time needed for transportation from place Q_j to terminal T_l is x_{jl}, then the time needed for transportation from P_i to D_k via Q_j by use of terminal T_l is equal to $a_{ij} + x_{jl} + c_{lk}$.

Suppose that the time which passengers traveling from place P_i to destination D_k have available for the transportation is denoted by b_{ik}. Our task is to choose the appropriate times $x_{jl},\ j \in N, l \in S$ such that the time for which the passengers from place P_i reach the destination D_k is equal to a given number b_{ik} for all $i \in M$ and for all $k \in R$, i.e.,

$$\max_{j \in N,\, l \in S} \{a_{ij} + x_{jl} + c_{lk}\} = b_{ik}. \tag{1}$$

2 Preliminaries

Max-plus algebra is the triple $(\overline{\mathbb{R}}, \oplus, \otimes)$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$ and $\oplus$, $\otimes$ are binary operations defined as follows:

$$a \oplus b = \max\{a, b\} \quad \text{and} \quad a \otimes b = a + b.$$

[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

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 consider the *ordering* $\leq$ on the sets $\overline{\mathbb{R}}(m,n)$ and $\overline{\mathbb{R}}(n)$ defined as follows:

- for $A, C \in \overline{\mathbb{R}}(m,n):\ A \leq C$ if $a_{ij} \leq c_{ij}$ for each $i \in M$ and for each $j \in N$,
- for $x, y \in \overline{\mathbb{R}}(n):\ x \leq y$ if $x_j \leq y_j$ for each $j \in N$.

We will use the *monotonicity of* $\otimes$, which means that for each $A, C \in \overline{\mathbb{R}}(m,n)$ and for each $B,\ D \in \overline{\mathbb{R}}(n,s)$ the implication

$$\text{if } A \leq C \text{ and } B \leq D \text{ then } A \otimes B \leq C \otimes D$$

holds.

3 Matrix equations

First, we define a system of max-plus linear equations. For $A \in \overline{\mathbb{R}}(m,n)$ and $b \in \overline{\mathbb{R}}(m)$ we can write a *system of max-plus linear equations* in the matrix form

$$A \otimes x = b. \tag{2}$$

It is well known (see [2, 14]) that system (2) is solvable if and only if the vector $x^*(A,b)$, defined by

$$x_j^*(A,b) = \min_{i \in M}\{b_i - a_{ij}\} \tag{3}$$

for any $j \in N$, is its solution. The vector $x^*(A,b)$ is called a *principal solution* of system (2).

The main topic of this article is solving matrix equations of the form

$$A \otimes X \otimes C = B \tag{4}$$

where $A \in \overline{\mathbb{R}}(m,n)$, $B \in \overline{\mathbb{R}}(m,r)$, and $C \in \overline{\mathbb{R}}(s,r)$ are given matrices and we are looking for an unknown matrix $X \in \overline{\mathbb{R}}(n,s)$.

Returning to Example 1, equality (1) can be written in max-plus algebra in the form $[A \otimes X \otimes C]_{ik} = b_{ik}$. Therefore, the solvability of the problem in mentioned example is equivalent to the solvability of (4). We will suppose that

- $b_{ik} \in \mathbb{R}$ for each $i \in M$ and for each $k \in R$;
- for each $j \in N$ there exists $i \in M$ such that $a_{ij} \in \mathbb{R}$;
- for each $l \in S$ there exists $k \in R$ such that $c_{lk} \in \mathbb{R}$.

In terms of our example, we suppose that

- from any place P_i to any destination D_k travel any passengers;
- through each check point Q_j pass passengers from at least one of places P_i;
- each terminal T_k handles passengers to at least one destination D_k.

In [10], there were defined so-called *principal matrix solution* $X^*(A,B,C) \in \overline{\mathbb{R}}(n,s)$ as follows:

$$x_{jl}^*(A,B,C) = \min_{k \in R}\{x_j^*(A,B_k) - c_{lk}\} \tag{5}$$

for each $j \in N$ and for each $l \in S$, where B_k is k-th column of B. The above three conditions ensure the finiteness of $X^*(A,B,C)$. The following theorem expresses the importance of the principal matrix solution for the solvability of (4).

Theorem 1 (Myšková [10]). *Let $A \in \overline{\mathbb{R}}(m,n)$, $B \in \overline{\mathbb{R}}(m,r)$ and $C \in \overline{\mathbb{R}}(m,n)$.*
a) *If $A \otimes X \otimes C = B$ for $X \in \overline{\mathbb{R}}(n,s)$, then $X \leq X^*(A,B,C)$;*
b) *$A \otimes X^*(A,B,C) \otimes C \leq B$;*
c) *The matrix equation $A \otimes X \otimes C = B$ is solvable if and only if the matrix $X^*(A,B,C)$ is its solution.*

Lemma 2 (Myšková [10]). Let $A,\ A^{(1)}, A^{(2)} \in \overline{\mathbb{R}}(m,n)$, $B,\ B^{(1)}, B^{(2)} \in \overline{\mathbb{R}}(m,r)$, and $C,\ C^{(1)},\ C^{(2)} \in \overline{\mathbb{R}}(s,r)$. The following assertions hold:
a) if $A^{(2)} \leq A^{(1)}$ then $X^*(A^{(1)},B,C) \leq X^*(A^{(2)},B,C)$;
b) if $B^{(1)} \leq B^{(2)}$ then $X^*(A,B^{(1)},C) \leq X^*(A,B^{(2)},C)$;
c) if $C^{(2)} \leq C^{(1)}$ then $X^*(A,B,C^{(1)}) \leq X^*(A,B,C^{(2)})$.

Another possibility for solving a matrix equation is using a *tensor product*.

Definition 1. Let $A = (a_{ij})$ be an $m \times n$ matrix and let $B = (b_{ij})$ be an $r \times s$ matrix. The *tensor product* of A and B is the following $mr \times ns$ matrix:

$$A \boxtimes B = \begin{pmatrix} A \otimes b_{11} & A \otimes b_{12} & \dots & A \otimes b_{1s} \\ A \otimes b_{21} & A \otimes b_{22} & \dots & A \otimes b_{2s} \\ \dots & \dots & \dots & \dots \\ A \otimes b_{r1} & A \otimes b_{r2} & \dots & A \otimes b_{rs} \end{pmatrix}.$$

Let $X \in B(n, s)$. Denote by $\operatorname{vec}(X)$ the vector $(X_1, X_2, \dots, X_s)^\top$, where X_l is l-th column of matrix X. Similarly we define $\operatorname{vec}(B)$.

Theorem 3 (Butkovič–Fiedler [1])**.** *Matrix equation*

$$(A_1 \otimes X \otimes C_1) \oplus (A_2 \otimes X \otimes C_2) \oplus \dots \oplus (A_r \otimes X \otimes C_r) = B, \tag{6}$$

where A_i, C_i, and B are matrices of compatible sizes, is equivalent to the vector-matrix system

$$(A_1 \boxtimes C_1^\top \oplus A_2 \boxtimes C_2^\top \oplus \dots A_r \boxtimes C_r^\top) \otimes \operatorname{vec}(X) = \operatorname{vec}(B). \tag{7}$$

Corollary 4. A matrix equation (4) is solvable if and only if a system of max-plus linear equations of the form

$$(A \boxtimes C^\top) \otimes \operatorname{vec}(X) = \operatorname{vec}(B) \tag{8}$$

is solvable.

4 Interval case

A certain disadvantage of the necessary and sufficient condition for the solvability of (4) given in Theorem 1 (iii) stems from the fact that it only indicates the existence or non-existence of the solution but does not indicate any action to be taken to increase the degree of solvability. However, it happens quite often in modelling 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. In practice, the traveling times a_{ij}, c_{lk} may depend on external conditions, so they are from intervals of possible values. Due to this fact, we will require the transportation times from P_i to D_k to be from a given intervals of possible values.

4.1 Interval systems of linear equations

Similarly to [4, 5, 13] we define an *interval matrix* $\boldsymbol{A}$ and *interval vector* $\boldsymbol{b}$ as follows:

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \overline{\mathbb{R}}(m,n);\ \underline{A} \le A \le \overline{A} \right\}, \quad \boldsymbol{b} = [\underline{b}, \overline{b}] = \left\{ b \in \overline{\mathbb{R}}(m);\ \underline{b} \le b \le \overline{b} \right\},$$

where $\underline{A}, \overline{A} \in \overline{\mathbb{R}}(m,n)$, $\underline{A} \le \overline{A}$ and $\underline{b}, \overline{b} \in \overline{\mathbb{R}}(m)$, $\underline{b} \le \overline{b}$.

Denote by

$$\boldsymbol{A} \otimes x = \boldsymbol{b} \tag{9}$$

the set of all of max-plus systems of linear equations of the form (2) such that $A \in \boldsymbol{A},\ b \in \boldsymbol{b}$. We shall call (9) a *max-plus interval system of linear equations*.

Definition 2. Interval system (9) is *T4 solvable* if there exist $b \in \boldsymbol{b}$ and $x \in \overline{\mathbb{R}}(n)$ such that $A \otimes x = b$ for each $A \in \boldsymbol{A}$.

The T4 solvability of interval system (9) was studied in [6]. We will just give you the essential assertions required in the next section. We define the *T4 sequence* of (9) as the sequence $\{b^{(t)}\}_{t=0}^{\infty}$ given by

$$b^{(t)} = \begin{cases} \overline{b} & \text{for } t = 0, \\ \underline{A} \otimes x^*(\overline{A}, b^{(t-1)}) & \text{for } t \ge 1. \end{cases} \tag{10}$$

Lemma 5 (Myšková [6])**.** If $b^{(t+1)} \ne b^{(t)}$ for each $t \in \mathbb{N}_0$ then there exists $r \in M$ such that the sequence $\{b_r^{(t)}\}_{t=0}^{\infty}$ is almost linear periodic with linear factor $q \in \mathbb{R}$, $-\infty \ne q < 0$, i. e., there are $p \in \mathbb{N}$ and $d \in \mathbb{N}_0$ such that for every $T > d$

$$b_r^{(T+p)} = b_r^{(T)} + p \times q \tag{11}$$

where $+$ and $\times$ are classical addition and multiplication, respectively.

Theorem 6 (Myšková [6])**.** *Interval system* (9) *is T4 solvable if and only if there exists $l \in \mathbb{N}_0$ such that $b^{(l+1)} = b^{(l)}$ and $b^{(l)} \in \boldsymbol{b}$.*

4.2 Interval matrix equations

We define

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \overline{\mathbb{R}}(m,n);\ \underline{A} \leq A \leq \overline{A} \right\}, \quad \boldsymbol{B} = [\underline{B}, \overline{B}] = \left\{ B \in \overline{\mathbb{R}}(m,r);\ \underline{B} \leq B \leq \overline{B} \right\},$$

and

$$\boldsymbol{C} = [\underline{C}, \overline{C}] = \left\{ C \in \overline{\mathbb{R}}(s,r);\ \underline{C} \leq C \leq \overline{C} \right\}.$$

Denote by

$$\boldsymbol{A} \otimes X \otimes \boldsymbol{C} = \boldsymbol{B} \tag{12}$$

the set of all matrix equations of the form (4) such that $A \in \boldsymbol{A},\ B \in \boldsymbol{B}$ and $C \in \boldsymbol{C}$. We call equation (12) an *interval max-plus matrix equation*.

We shall think over the solvability of interval max-plus matrix equation on the ground of solvability of matrix equations of the form (4) where $A \in \boldsymbol{A},\ B \in \boldsymbol{B}$, and $C \in \boldsymbol{C}$. We can define several types of solvability of interval max-plus matrix equations. Some of them (see Table 1) have been studied in [10, 11], where you can find necessary and sufficient conditions for them.

Type of solvability	**Condition**
tolerance	$(\exists X \in \overline{\mathbb{R}}(n,s))(\forall A \in \boldsymbol{A})(\forall C \in \boldsymbol{C})(\exists B \in \boldsymbol{B});\ A \otimes X \otimes C = B$
right-weakly tolerance	$(\forall C \in \boldsymbol{C})(\exists X \in \overline{\mathbb{R}}(n,s))(\forall A \in \boldsymbol{A})(\exists B \in \boldsymbol{B});\ A \otimes X \otimes C = B$
left-weakly tolerance	$(\forall A \in \boldsymbol{A})(\exists X \in \overline{\mathbb{R}}(n,s))(\forall C \in \boldsymbol{C})(\exists B \in \boldsymbol{B});\ A \otimes X \otimes C = B$
weakly tolerance	$(\forall A \in \boldsymbol{A})(\forall C \in \boldsymbol{C})(\exists X \in \overline{\mathbb{R}}(n,s))(\exists B \in \boldsymbol{B});\ A \otimes X \otimes C = B$
strongly universal	$(\exists X \in \overline{\mathbb{R}}(n,s))(\forall A \in \boldsymbol{A})(\forall C \in \boldsymbol{C})(\forall B \in \boldsymbol{B});\ A \otimes X \otimes C = B$
universal	$(\forall B \in \boldsymbol{B})(\exists X \in \overline{\mathbb{R}}(n,s))(\forall A \in \boldsymbol{A})(\forall C \in \boldsymbol{C});\ A \otimes X \otimes C = B$
weakly universal	$(\forall A \in \boldsymbol{A})(\forall C \in \boldsymbol{C})(\forall B \in \boldsymbol{B})(\exists X \in \overline{\mathbb{R}}(n,s));\ A \otimes X \otimes C = B$

Table 1 Types of solvability of (12)

Strongly tolerance solvability

In this part, we shall deal with other solvability concept, the strongly tolerance solvability.

Definition 3.
a) A matrix $B \in \boldsymbol{B}$ is a *strongly tolerance matrix* of interval max-plus matrix equation (12) if there exists a matrix $X \in B(n,s)$ such that $A \otimes X \otimes C = B$ for each $A \in \boldsymbol{A}$ and for each $C \in \boldsymbol{C}$.
b) Interval max-plus matrix equation (12) is *strongly tolerance solvable* if there exists a matrix $B \in \boldsymbol{B}$ such that B is a strongly tolerance matrix of (12).

Theorem 7 (Myšková [11])**.** *Interval max-plus matrix equation* (12) *is strongly universally solvable if and only if*

$$\underline{B} = \overline{B} = B \quad \textit{and} \quad \underline{A} \otimes X^*(\overline{A}, B, \overline{C}) \otimes \underline{C} = B.$$

Lemma 8. A matrix $B \in \boldsymbol{B}$ is a strongly tolerance matrix of interval max-plus matrix equation (12) if and only if it satisfies equality

$$\underline{A} \otimes X^*(\overline{A}, B, \overline{C}) \otimes \underline{C} = B. \tag{13}$$

Proof. A matrix $B \in \boldsymbol{B}$ is a strongly tolerance matrix of (12) if and only if interval max-plus matrix equation (12) with the constant right-hand side $\underline{B} = \overline{B} = B$ is strongly universally solvable, which is according to Theorem 7 equivalent to (13). □

We define the sequence of matrices $\{B^{(t)}\}_{t=0}^{\infty}$ as follows:

$$B^{(t)} = \begin{cases} \overline{B} & \text{for } t = 0, \\ \underline{A} \otimes X^*(\overline{A}, B^{(t-1)}, \overline{C}) \otimes \underline{C} & \text{for } t \geq 1. \end{cases} \tag{14}$$

Lemma 9. The sequence $\{B^{(t)}\}_{t=0}^{\infty}$ defined by (14) is non-increasing.

Proof. By Theorem 1 b) and by the monotonicity of $\otimes$ we have

$$B^{(t+1)} = \underline{A} \otimes X^*(\overline{A}, B^{(t)}, \overline{C}) \otimes \underline{C} \leq \overline{A} \otimes X^*(\overline{A}, B^{(t)}, \overline{C}) \otimes \overline{C} \leq B^{(t)}.$$

□

Lemma 10. Let $B \in \boldsymbol{B}$ be a strongly tolerance matrix of interval max-plus matrix equation (12). Then the inequality $B \leq B^{(t)}$ is satisfied for each $t \in \mathbb{N}_0$.

Proof. By mathematical induction on t
1. For $t = 0$ the inequality $B \leq \overline{B} = B^{(0)}$ trivially holds.
2. Suppose that $B \leq B^{(t)}$. We obtain $B = \underline{A} \otimes X^*(\overline{A}, B, \overline{C}) \otimes \underline{C} \leq \underline{A} \otimes X^*(\overline{A}, B^{(t)}, \overline{C}) \otimes \underline{C} = B^{(t+1)}$.

□

Lemma 11. If there is no $t \in \mathbb{N}$ such that $B^{(t+1)} = B^{(t)}$ then there exist $i \in M$, $k \in R$, $q \in \mathbb{R}$, $-\infty \neq q < 0$, $p \in \mathbb{N}$ and $d \in \mathbb{N}_0$ such that for every $T > d$

$$b_{ik}^{(T+p)} = b_{ik}^{(T)} + p \times q. \tag{15}$$

Proof. Let us take the sequence $\{\mathrm{vec}\,(B^{(t)})\}_{t=0}^{\infty}$. Using the definition of a tensor product and Corollary 4 we can show that the sequence $\{\mathrm{vec}\,(B^{(t)})\}_{t=0}^{\infty}$ is equal to the T4 sequence $\{d^{(t)}\}_{t=0}^{\infty}$ of the interval system of linear equations of the form $\boldsymbol{E} \otimes \mathrm{vec}\,(X) = \boldsymbol{d}$, where $\boldsymbol{E} = [\underline{A} \boxtimes \underline{C}^{\top}, \overline{A} \boxtimes \overline{C}^{\top}]$ and $\boldsymbol{d} = [\mathrm{vec}\,(\underline{B}), \mathrm{vec}\,(\overline{B})]$. According to Lemma 5 there exists $v \in \{1, 2, \ldots, mr\}$, $-\infty \neq q < 0$, $p \in \mathbb{N}$ and $d \in \mathbb{N}_0$ such that for every $T > d$ we have $d_v^{(T+p)} = d_v^{(T)} + p \times q$. This is equivalent to the existence of $i \in M$, $k \in R$, $-\infty \neq q < 0$, $p \in \mathbb{N}$ and $d \in \mathbb{N}$ such that for every $T > d$ equation (15) holds. □

Theorem 12. *Interval max-plus matrix equation* (12) *is strongly tolerance solvable if and only if there exists* $u \in \mathbb{N}_0$ *such that* $B^{(u+1)} = B^{(u)}$ *and* $B^{(u)} \in \boldsymbol{B}$.

Proof. If $B^{(u+1)} = B^{(u)}$ and $B^{(u)} \in \boldsymbol{B}$ for some $u \in \mathbb{N}_0$ then $\underline{A} \otimes X^*(\overline{A}, B^{(u)}, \overline{C}) \otimes \underline{C} = B^{(u)}$ which means that matrix $B^{(u)}$ is a strongly tolerance matrix of (12). Thus interval max-plus matrix equation (12) is strongly tolerance solvable.

For the converse implication, for the sake of contradiction suppose that there is no $u \in \mathbb{N}_0$ such that $B^{(u+1)} = B^{(u)}$, where $B^{(u)} \in \boldsymbol{B}$ and (12) is strongly tolerance solvable with a strongly tolerance matrix $B \in \boldsymbol{B}$. There are two possibilities: either $B^{(u+1)} \neq B^{(u)}$ for each $u \in \mathbb{N}_0$ or there exists $u \in \mathbb{N}_0$ such that $B^{(u+1)} = B^{(u)}$, but $B^{(u)} \notin \boldsymbol{B}$.

In the first case, according to Lemma 11 there exists $i \in M$, $k \in R$, $q \in \mathbb{R}$, $-\infty \neq q < 0$, $p \in \mathbb{N}$ and $d \in \mathbb{N}_0$ such that for every $T > d$ the equality $b_{ik}^{(T+p)} = b_{ik}^{(T)} + p \times q$ holds. By Lemma 10 there exists $t^* \in \mathbb{N}_0$ such that $b_{ik}^{(t^*)} < \underline{b}_{ik}$. Then $b_{ik} \leq b_{ik}^{(t^*)} < \underline{b}_{ik}$, i. e., $B \notin \boldsymbol{B}$, a contradiction.

In the second case, suppose that there exists $u \in \mathbb{N}_0$ such that $B^{(u+1)} = B^{(u)}$, $B^{(u)} \notin \boldsymbol{B}$ and there exists a matrix $B \in \boldsymbol{B}$ such that B is the strongly tolerance matrix. The definition of the sequence $\{B^{(t)}\}_{t=0}^{\infty}$ and Lemma 9 imply that $\underline{B} \not\leq B^{(u)}$. This means that there exist $i \in M$, $k \in R$ such that $b_{ik}^{(u)} < \underline{b}_{ik}$. By Lemma 10 we get $b_{ik} \leq b_{ik}^{(u)} < \underline{b}_{ik}$ and consequently $B \notin \boldsymbol{B}$, a contradiction. □

Theorem 12 enables us to give the following algorithm for checking the strongly tolerance solvability.

Algorithm ST
Input: $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$.

Output: 'yes' in variable st and strongly tolerance matrix B^* if the given interval max-plus matrix equation is strongly tolerance solvable; 'no' in st otherwise.

begin
(i) $B^{(0)} = \overline{B}$, $t = 0$;
(ii) $B^{(t+1)} = \underline{A} \otimes X^*(\overline{A}, B^{(t)}, \overline{C}) \otimes \underline{C}$;
(iii)**If** $\underline{B} \not\leq B^{(t+1)}$ **then** st :='no', go to **end**;
(iv)**If** $B^{(t+1)} = B^{(t)}$ **then** st :='yes', $B^* = B^{(t)}$, **go to end**;
(v) $t := t + 1$, **go to** (ii);

end

Example 2. Let us take

$$\boldsymbol{A}=\begin{pmatrix} [1,3] & [4,4] & [4,4] \\ [3,4] & [3,4] & [2,2] \\ [2,4] & [4,5] & [3,3] \end{pmatrix},\ \boldsymbol{B}=\begin{pmatrix} [14,19] & [11,13] & [15,20] \\ [12,19] & [11,17] & [12,17] \\ [13,18] & [12,16] & [14,19] \end{pmatrix},\ \text{and}\ \boldsymbol{C}=\begin{pmatrix} [4,4] & [1,2] & [2,3] \\ [3,3] & [4,5] & [1,4] \\ [1,3] & [1,1] & [4,4] \end{pmatrix}.$$

We check whether the interval matrix equation $\boldsymbol{A}\otimes X\otimes\boldsymbol{C}=\boldsymbol{B}$ is strongly tolerance solvable.

Solution. We compute $\{B^{(t)}\}_{t=0}^{\infty}$. We obtain the sequence

$$\left\{\begin{pmatrix} 19 & 13 & 20 \\ 19 & 17 & 17 \\ 18 & 16 & 19 \end{pmatrix},\begin{pmatrix} 15 & 13 & 16 \\ 15 & 13 & 16 \\ 15 & 13 & 16 \end{pmatrix},\begin{pmatrix} 15 & 13 & 16 \\ 14 & 12 & 15 \\ 14 & 12 & 15 \end{pmatrix},\begin{pmatrix} 15 & 13 & 16 \\ 13 & 11 & 14 \\ 14 & 12 & 15 \end{pmatrix},\begin{pmatrix} 15 & 13 & 16 \\ 13 & 11 & 14 \\ 14 & 12 & 15 \end{pmatrix},\dots\right\}.$$

Since $B^{(4)}=B^{(3)}\geq\underline{B}$, according to Theorem 12 the given interval max-plus matrix equation is strongly tolerance solvable with the strongly tolerance matrix $B^*=B^{(3)}$. Moreover, we can compute the matrix X^* such that $A\otimes X^*\otimes C=B^*$ for each $A\in\boldsymbol{A}$ and for each $C\in\boldsymbol{C}$. We obtain

$$X^*=X^*(\overline{A},B^*,\overline{C})=\begin{pmatrix} 5 & 2 & 6 \\ 5 & 2 & 6 \\ 7 & 4 & 8 \end{pmatrix}.$$

4.3 Conclusion

In this paper, we dealt with the solvability of interval matrix equations in max-plus algebra. Max-plus 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 transportation times. In economics, those values can represent the production costs of each stage of product production. Suppose that we can not influence production costs a_{ij} and c_{lk}, i.e., we only know the intervals at which these values are. Using Algorithm ST we can find the total production costs and corresponding values x_{jl}, selecting which we achieve these total production costs independently on values a_{ij} and c_{lk}.

References

[1] Butkovič., P. & Fiedler., M. Tropical tensor product and beyond. `http://web.mat.bham.ac.uk/P.Butkovic/Pubs.html`.

[2] Cuninghame-Green, R. A. (1979). *Minimax Algebra*. Lecture notes in Economics and Mathematical systems. Berlin: Springer.

[3] Gavalec, M. & Plavka, J. (2010). Monotone interval eigenproblem in max-plus algebra. *Kybernetika*, 46(3), 387–396.

[4] Myšková, H. (2005). Interval systems of max-separable linear equations. *Lin. Algebra Appl.*, 403, 263–272.

[5] Myšková, H. (2006). Control solvability of interval systems of max-separable linear equations. *Lin. Algebra Appl.*, 416, 215–223.

[6] Myšková, H. (2012). On an algorithm for testing T4 solvability of fuzzy interval systems. *Kybernetika*, 48(5), 924–938.

[7] Myšková, H. (2012). Robustness of interval Toeplitz matrices in fuzzy algebra. *Acta Electrotechnica et Informatica*, 12 (4), 56–60.

[8] Myšková, H. (2012). Interval eigenvectors of circulant matrices in fuzzy algebra. *Acta Electrotechnica et Informatica* 12 (3), 57–61.

[9] Myšková, H. (2012). Weak stability of interval orbits of circulant matrices in fuzzy algebra. *Acta Electrotechnica et Informatica* 12 (3), 51–56.

[10] Myšková, H. (2016). Interval max-plus matrix equations. *Lin. Algebra Appl.*, 492, 111–127.

[11] Myšková, H. (2018). Universal solvability of interval max-plus matrix equations. *Discrete Appl. Math.*, 239, 165–173.

[12] Olsder, G. J. et al. (1998). Course Notes: Max-algebra Aproach to Discrete Event Systems. *Algebres Max-Plus et Applications an Informatique et Automatique*, INRIA, 1998, 147–196.

[13] Plavka, J. (2012): On the $O(n^3)$ algorithm for checking the strong robustness of interval fuzzy matrices. *Discrete Appl. Math.* 160, 640–647.

[14] Zimmermann, K. (1976). *Extremální algebra*. Praha: Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV.

Nonparametric prediction of indices from the Central European stock exchanges

Witold Orzeszko[1], Sylwester Bejger[2]

Abstract. Nonparametric regression is an alternative to the parametric approach, where parametric models, i.e. models of the certain functional form with a fixed number of parameters, are applied. As opposed to the parametric approach, nonparametric regression models do not impose restrictive assumptions about the form of the modeling dependencies and in consequence, they are more flexible and let the data speak for themselves. That is why they are a promising tool for forecasting, especially in a case of nonlinear time series.
One of the most popular nonparametric regression method is kernel smoothing. The aim of the paper is to assess the usefulness of the selected kernel smoothers to predict log returns of the indices from the Polish, Czech and Hungarian stock markets. Two kernel smoothers were applied in the research. The first one was the Nadaraya-Watson estimator, which is the most popular nonparametric regression method. The second one was the local-linear kernel estimator, which combines the local linear approximation and the kernel smoother. For comparison, the linear regression model and the naive method were also considered. Lagged by one day: the S&P 500 Index, the currencies exchange rates, and the autoregressive variable were used as the predictors in the applied regression models.

Keywords: kernel smoothers, nonparametric prediction, stock indices, Nadaraya-Watson estimator, local-linear kernel estimator.

JEL Classification: C14, C53, C58
AMS Classification: 62G08, 62J05, 62P20

1 Kernel smoothing

One of the major tasks of financial econometrics is to identify regularities occurring in financial data. Discovering the mechanisms of shaping real processes in the financial markets is essential not only for cognitive, but also for the applicational reasons, since it forms the basis of their modeling and forecasting. Identifying relationships amongst financial processes is part of the current research conducted on financial markets efficiency – the concept whose significance for the development of finance was confirmed by awarding the 2013 Nobel Prize in Economic Sciences to its pioneer - Eugene Fama. Among the tools applied for analyzing financial processes, nonparametric methods gain increasing recognition. These methods do not impose restrictive assumptions about the form of the modeling dependencies and in consequence, they are more flexible and let the data speak for themselves. That is why they can be very useful in data forecasting, especially in case of nonlinear time series.

The main tools for forecasting real data are regression models. Generally, these models can be written as

$$Y = m(\boldsymbol{x}) + \varepsilon \tag{1}$$

where $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$, Y is the dependent variable, $\mathbf{X}$ – the vector of regressors, $\mathbf{x}$ – some fixed value of $\mathbf{X}$, and the error term ε has the properties $E(\varepsilon|\mathbf{X} = \mathbf{x}) = 0$ and $E(\varepsilon^2|\mathbf{X} = \mathbf{x}) = \sigma^2(\mathbf{x})$ (Pagan & Ullah [5]). In the parametric approach, the function $m(\mathbf{x})$ is chosen from some parametric class so that the specific candidate is obtained by specifying a fixed finite number of parameters. In nonparametric regression any certain parametric form for $m(\mathbf{x})$ is assumed. Instead of it, $m(\mathbf{x})$ is approximated by a function from some flexible class, in such a way that the approximation precision increases with the sample size (cf. Härdle, Lütkepohl & Chen [3])

One of the most important nonparametric regression methods are kernel smoothers. The Nadaraya-Watson estimator (N-W hereafter, Nadaraya [4], Watson [8]) is the most popular example of such a smoother. In the univariate case it is defined as

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)}, \tag{2}$$

[1] Nicolaus Copernicus University, ul. Gagarina 13a, Toruń, Poland, e-mail: witold.orzeszko@umk.pl.
[2] Nicolaus Copernicus University, ul. Gagarina 13a, Toruń, Poland, e-mail: sylwester.bejger@umk.pl.

where n is a sample size, K is the kernel function and h is the bandwidth. The kernel is a nonnegative, even function which satisfies

$$\int_{-\infty}^{+\infty} K(x)dx = 1, \tag{3}$$

$$\int_{-\infty}^{+\infty} xK(x)dx = 0. \tag{4}$$

In practice, it is also expected that K has the global maximum in $x_0 = 0$, which assures that the estimated value $\hat{m}(x)$ mainly depends on observations y_i corresponding to x_i which are close to x. Commonly used kernel functions include the Gaussian, Epanechnikov, triangle and biweight.

In the multivariate case, to calculate $\hat{m}(\mathbf{x})$ for $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ the formula (2) is applied, but the function $K(x)$ is replaced by the multivariate kernel function $\kappa(\mathbf{x})$. In practice, the most popular forms of $\kappa(\mathbf{x})$ are the product kernel

$$\kappa(\boldsymbol{x}) = \prod_{i=1}^{d} K(x_i) \tag{5}$$

and the radial kernel

$$\kappa(\boldsymbol{x}) = c \cdot K\left(\sqrt{\sum_{i=1}^{d} x_i^2}\right). \tag{6}$$

The local-linear kernel estimator (LLKE hereafter) is a combination of the local linear approximation and the kernel smoother. This method estimates the regression function $\hat{m}(x)$ by "locally" fitting the first degree polynomial to the data via weighted least squares with weights defined by the kernel function (cf. Stone [7], Fan & Gijbels [2]). It consists on finding the parameters $\hat{\beta}_0, \hat{\beta}_1$ of the approximating polynomial that minimize the formula

$$\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)\left(y_i - \beta_0 - \beta_1(x_i - x)\right)^2. \tag{7}$$

The solution of this minimization problem is

$$\hat{m}(x) = \sum_{i=1}^{n} w_i(x) y_i, \tag{8}$$

where

$$w_i(x) = \frac{K\left(\frac{x-x_i}{h}\right)\left(\hat{s}_2(x) - (x-x_i)\hat{s}_1(x)\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)\left(\hat{s}_2(x) - (x-x_i)\hat{s}_1(x)\right)}, \tag{9}$$

and

$$\hat{s}_r(x) = \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)(x - x_i)^r, \text{ for } r = 1, 2. \tag{10}$$

The local-linear kernel estimator is more general to the Nadaraya-Watson estimator since N-W fits a constant to the data close to x whereas LLKE fits a straight line.

2 Application of kernel smoothers to predict indices from the Central European stock exchanges

2.1 Investigated data, applied methods and verification techniques

The aim of the research is to assess the usefulness of the both kernel smoothers presented in Section 1 to predict the main indices from the Polish, Czech and Hungarian stock markets, i.e. respectively, WIG, PX and BUX indices. The daily log returns from January 2, 2015 to March 29, 2018 were predicted. In the regression models the forecasts y_T^* were calculated from the formula $y_T^* = \hat{m}(x_{T-1})$, where the regression function m was estimated for each T, based on the previous 1000 observations. Log returns of the S&P 500 Index, the local currencies exchange rates against the US dollar (respectively: USD/PLN, USD/CZK, USD/HUF) and the autoregressive variable were used as the predictors x_T. For each investigated index three regression models with one single predictor were considered, but additionally the models with all three variables were also applied.

The Gaussian kernel function, given by

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \tag{11}$$

was applied in calculations, and the product kernel was used in the three-variate models. The bandwidth parameter h was determined using the cross-validation technique.

For comparison, the linear regression model and the naïve method were also considered. In the naïve method

$$y_T^* = \frac{\sum_{i=1}^{1000} y_{T-i}}{1000}, \tag{12}$$

which means that the predicted value is the average of the previous 1000 observations.

To assess the forecasting abilities of the analyzed methods, two criteria of forecasts quality were applied. The first one was the Root Mean Square Error (RMSE), given by

$$RMSE = \sqrt{\frac{1}{h}\sum_{t=1}^{h}(y_{n+t} - y_{n+t}^*)^2}, \tag{13}$$

where h is the number of forecasts.

The second applied criterion was the Diebold-Mariano test for predictive accuracy (Diebold & Mariano [1]). Let y_{T1}^* and y_{T2}^* denote forecasts obtained from two competitive models, $e_{T1} = y_T - y_{T1}^*$, $e_{T2} = y_T - y_{T2}^*$ - forecasts errors, and $d_T = e_{T1}^2 - e_{T2}^2$ - the loss function. The null hypothesis (H_0: $E(d) = 0$) states that the forecasts from both models have the same accuracy and the alternative hypothesis (H_1: $E(d) \neq 0$) – that they have different levels of accuracy. The test statistic $DM = \frac{\bar{d}}{\sqrt{\frac{var(d)}{h}}}$ (where $\bar{d}$ and $var(d)$ are, respectively, the mean and the variance of the loss function) is asymptotically normally distributed under the null.

2.2 Results of the research

First, the RMSE errors were calculated. The calculated values for the naïve method are presented in Table 1.

WIG	PX	BUX
0.00896	0.00858	0.01030

Table 1 RMSE errors for the naïve method

For better clarity, the RMSE errors for other forecasting methods were transformed to the relative errors, according to the formula $\frac{RMSE}{RMSE_{naive}} \cdot 100\%$ (see Tables 2-4). The values smaller than 100% (indicating the superiority over the naïve method) were bolded.

linear	N-W	LLKE
Autoregressive		
99.77%	**99.89%**	**99.92%**
S&P 500		
96.74%	**97.51%**	**96.52%**
USD/PLN		
99.64%	**99.94%**	**99.76%**
Three variables		
96.84%	**97.51%**	**97.04%**

Table 2 Relative RMSE errors for the WIG index

linear	N-W	LLKE
Autoregressive		
100.30%	100.02%	101.39%
S&P 500		
97.43%	**99.14%**	100.19%

	USD/CZK	
100.11%	**99.89%**	100.24%
	Three variables	
97.12%	**97.99%**	**98.05%**

Table 3 RMSE errors for the PX index

linear	**N-W**	**LLKE**
	Autoregressive	
100.12%	100.13%	100.21%
	S&P 500	
99.76%	100.93%	100.80%
	USD/HUF	
100.00%	100.36%	106.96%
	Three variables	
99.97%	100.97%	100.63%

Table 4 RMSE errors for the BUX index

As it can be seen, the most promising results were obtained for the WIG index. In this case, all of the calculated relative RMSE errors are smaller than 100%, which indicates that the regression models give more accurate forecasts than the naïve method. The best results were obtained from the model with the predictor S&P500 and from the model with all three predictors. These models led to the most accurate forecasts also in the case of the two remaining indices. However, for the BUX index, utility of these models is not convincing.

Moreover, we can conclude that the superiority of the kernel smoothers over the linear models seems rather exceptional. Furthermore, when comparing both kernel smoothers, the local-linear kernel estimator usually did not lead to better results than the Nadaraya-Watson estimator.

Next, the Diebold-Mariano test was performed (see Tables 5-7). In each cell, one can see the value of the test statistic. The symbols *, ** denote rejection of H_0 at 5% and 1% significant level, respectively. However, besides significance, it is also interesting to analyze the sign of the calculated statistics. Their negative values indicate the superiority of the method in the stub of the table and its positive value - the superiority of the method in the box head.

	naïve	**linear**	**N-W**	**LLKE**
		Autoregressive		
linear	- 0.772	X	-0.503	-0.907
N-W	- 0.845	X	X	-0.136
LLKE	- 0.216	X	X	X
		S&P 500		
linear	-3.132**	X	-0.692	0.203
N-W	- 1.559	X	X	1.892
LLKE	-1.962*	X	X	X
		USD/PLN		
linear	- 1.129	X	-0.310	-1.382
N-W	- 1.207	X	X	-0.318
LLKE	- 0.784	X	X	X
		Three variables		
linear	-2.687**	X	-1.089	-1.749
N-W	-3.145**	X	X	0.729
LLKE	-2.449*	X	X	X

Table 5 Results of the Diebold-Mariano test for the WIG index

	naïve	linear	N-W	LLKE
Autoregressive				
linear	1.333	X	1.346	-1.007
N-W	0.299	X	X	-1.250
LLKE	1.255	X	X	X
S&P 500				
linear	-1.405	X	-1.339	-1.607
N-W	-0.442	X	X	-0.766
LLKE	0.065	X	X	X
USD/CZK				
linear	0.239	X	0.261	-0.962
N-W	0.159	X	X	-0.833
LLKE	0.594	X	X	X
Three variables				
linear	-1.679	X	-0.807	-0.902
N-W	-1.353	X	X	-0.045
LLKE	-0.792	X	X	X

Table 6 Results of the Diebold-Mariano test for the PX index

	naïve	linear	N-W	LLKE
Autoregressive				
linear	1.266	X	-0.028	-0.374
N-W	0.875	X	X	-0.525
LLKE	0.816	X	X	X
S&P 500				
linear	-0.445	X	-1.481	-0.950
N-W	0.769	X	X	0.284
LLKE	0.508	X	X	X
USD/HUF				
linear	-0.007	X	-1.283	-1.009
N-W	1.179	X	X	-0.956
LLKE	1.015	X	X	X
Three variables				
linear	-0.055	X	-1.265	-0.981
N-W	0.791	X	X	0.447
LLKE	0.611	X	X	X

Table 7 Results of the Diebold-Mariano test for the BUX index

The results of the Diebold-Mariano test indicate significant differences in predictive accuracy only in the case of the WIG index, when the predictor S&P 500 and the set of all three predictors were applied. In these both cases, the test showed that the regression models gave more accurate forecasts than the naïve method. No more significant differences were detected, neither for the other predictors, nor for the other indices. However, when analyzing the signs of the calculated statistics, one can conclude that it leads to the ranking of the applied methods which is generally consistent with the results of the RMSE errors (the slight differences may be observed only for the models with the exchange rates predictors).

3 Summary

The conducted research confirmed that stock returns inherently contain a sizable unpredictable component, so the best forecasting models can explain only a relatively small part of stock returns (cf. Rapach & Zhou [6]). Only for the WIG index the regression models revealed their superiority over the naïve method irrespective the applied predictor variables. However, one can conclude that the most accurate forecasts were obtained from the model with the predictor S&P500 and from the model with all three predictors. These predictors led to the best results in the case of the PX and BUX indices, as well.

Generally, the applied kernel smoothers did not confirm their superiority over the linear model. It may indicate that the mechanism which governs the dynamics of the stock prices is not nonlinear. When comparing both kernel smoothers, the local-linear kernel estimator, despite its more general form, usually did not lead to more accurate forecasts than the Nadaraya-Watson estimator.

References

[1] Diebold, F. X. & Mariano R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, *13*, 253–265.
[2] Fan, J. & Gijbels, I. (1992). Variable Bandwidth and Local Linear Regression Smoothers. *Annals of Statistics*, *20*, 2008–2036.
[3] Härdle, W., Lütkepohl, H. & Chen, R. (1997). A Review of Nonparametric Time Series Analysis. *International Statistical Review*, *65*, 49–72.
[4] Nadaraya, E. A. (1964). On Estimating Regression. *Theory of Probability and its Applications*, *9*, 141–142.
[5] Pagan, A. & Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge: Cambridge University Press.
[6] Rapach, D. & Zhou, G. (2013). Forecasting Stock Returns. In G. Elliot & A. Timmermann (Eds.), *Handbook of Economic Forecasting vol. 2A* (pp. 329–383). Amsterdam: Elsevier.
[7] Stone, C. J. (1977). Consistent Nonparametric Regression. *Annals of Statistics*, *5*, 595–620.
[8] Watson, G. S. (1964). Smooth Regression Analysis. *Sankhya: The Indian Journal of Statistics* (*Series A*), *26*, 359–372.

Main macroeconomic indicators in CR under exchange rate rule

Václava Pánková

Abstract The role of exchange rate and its possible regulation recently performed by the Czech National Bank is studied with the regard to the economic growth, exports and incoming FDI what are the macroeconomic entities stressed in the context of a positive evaluation of economic development in the CR.

Relevant economic theory is summarized and represented by econometric models with a multi-equational structure. Applying appropriate estimation techniques, the results show that the economic growth represented by GDP increments is supported by exports and exchange rate and also the regulation of the CZK/EUR parity evinces a positive effect. The incoming FDI do not affect significantly neither GDP nor exports. The amount of incoming FDI, as well as the export, evidently depends on the exchange rate but there is no apparent influence of its temporary regulation. For including out-coming FDI as an exogenous variable it also was found that FDI outcomes are substitutes of the exports in the Czech economic environment.

Keywords: econometric multi-equational models, econometric methods, exchange rate rule

JEL Classification: C3, E1
AMS Classification: 62J99

1 Introduction

For the small open economy of Czech Republic, GDP, the exports and incoming investments are of crucial importance. In the period from 2014 to 2017, CNB (Czech National Bank) regulated the CZK/EUR parity with the goal to support their growth or at least to prevent their decrease.

The flows of FDI are the fundamental elements in the economical evolution of countries within the globalization process of economy. The negative link between exchange rate volatility and FDI is known and as a special topic was treated e.g. by [2], [4]. For the economic policy makers it means to make an effort to maintain exchange rate stability in order to attract more FDI.

The economists have long known that the way how exchange rate is managed is important for economic growth. In [7] a more strictly formulation is presented: poorly managed exchange rates can be disastrous for economic growth. Exchange rate rule is one of the possibilities of macroeconomic targeting concept discussed with some important conclusions e.g. in [1] or [5]. In 2014 the CNB started to weaken the exchange rate of the Czech crown because of a disproportionately low inflation rate. This policy was dropped in April 2017 and there is no unified opinion about its suitability among the Czech economists.

The model studying a possible interaction among variables mentioned above is based on the relation between foreign trade and foreign direct investment (FDI) which is a subject of many theoretical as well as applied studies the representative survey of which is given e.g. in [9]. Empirical studies provide often contradictory evidence of the link between FDI and trade. Having FDI flowing from country A to country B three basic scenarios can occur

- Country A has a relative abundance of capital and country B has a relative abundance of labor
- Country B has a comparative advantage in specific production which is not related to the use of capital
- Country B has a sufficient and relative secure demand surplus.

Each scenario has a couple of sub-scenarios representing possible consequences and effect on export and import of both sides. That is why the output of a model is not to be anticipated.

In general, the model of the form $\Delta Y = \beta FDI + \gamma' X + u$ (see [9]) is used; Y is for export or import, X is a vector of further important variables. Regarding the fact that no long time-series of Czech annual data exist, X is specified

as containing increment of GDP, increment of exchange rate and dummy variable related to exchange rate intervention years.

The eventual influence of the exchange rate rule on the changes in export / import will then be seen straightforward. In case of its evidence, a possible influence on GDP (not quantified) implicitly follows from the identity comprising consumption, investment, governmental expenditure and balance of foreign trade. Of course, vice versa implication does not held; GDP may be influenced through some other summand(s).

As for FDI in the proposed model, the total value is a most rough figure. Basic categorization is to distinguish between FDI incoming and outcoming what helps to consider whether FDI complements or substitutes trade (detail e.g. in [6]). Trade and FDI are main strategies to access foreign markets. Two rival hypotheses exist concerning a relation between FDI and exports (i) Substitutes: Capital mobility represented by FDI removes differences in factor endowments, hence makes an international trade less advantageous. (ii) Complements: FDI create or expand the opportunity to export.

FDI may be refined according to a sector to which it is applied. In general, the database of Eurostat offers categories as agriculture, forestry, manufacture, finance, a.s.o. Unfortunately, the Czech Republic data is not so subtle detailed except of a very small couple of years. Though the category manufacture will probably be the most appropriate one, a very poor finding is that it is approximately one third of the total sum.

2 Data and Models

Employing data found in Eurostat concerning CR from 1997 to 2016 annually, the followed variables are GDP, exports EXP, imports IMP, foreign direct investment inflows FDIIN, foreign direct investment outflows FDIOUT, all measured in millions of Euro and exchange rate ER relating CZK to EUR.

Examining a possible non-stationarity, the results summarized in Table 1 were found.

variable	ADF-test	KPSS-test
GDP	I(1)	I(1)
EXP	I(1)	I(1)
IMP	I(1)	I(1)
ER	I(1)	I(1)
FDIIN	I(0)	I(0)
FDIOUT	I(0)	I(0)

Table 1 ADF/KPSS tests of stationarity

The cointegration was not studied because the data sample will not allow for using such tool as e.g. ECM model (unsufficient deg. of freedom). The increments of GDP, EXP, IMP, ER are used instead, described by the prefix D-, which as a consequence of the facts in Table 1 are stationary. DGDP, DEXP, DIMP are often used to represent actual economic growth. Variable DER respects the empirical knowledge of consequences of exchange rate volatility on important economic events. FDI´s are not differentiating what corresponds not only with the technical reasons but also with the idea of the models used to explain an impact of total FDI´s on the economic growth as well as their effect on trade flows. The philosophy of Model 1 is described in [9]. D is a dummy variable: D=0 in the years without an exchange rate intervention, D=1 covers the period of the exchange rate rule application.

Model 1

$$DEXP = \alpha_0 + \alpha_1 FDIIN + \alpha_2 FDIOUT + \alpha_3 DGDP + \alpha_4 DER + \alpha_5 D + u$$

$$DIMP = \beta_0 + \beta_1 FDIIN + \beta_2 FDIOUT + \beta_3 DGDP + \beta_4 DER + \beta_5 D + v$$

Estimating both equations by the SUR (seemingly unrelated regression) method, the results are summarized in Table 2. According to R-squared, the equations explain the endogenous variables well enough. The same cannot

be said about the unique variables. With the exception of constants, there is no evidence of any effect of exchange rate policy (parameters α_5, β_5) and probably surprisingly incoming FDI do not influence export. The computation was repeated with the above explained approximation FDImanufacture=FDIIN / 3 with no apparent changes.

System: MODEL 1
Estimation Method: Seemingly Unrelated Regression

	Coefficient	Std. Error	t-Statistic	Prob.
α_0	3385.202	3466.411	0.976573	0.3404
α_1	0.034881	0.652029	0.053496	0.9579
α_2	-2.063002	1.159958	-1.778514	0.0905
α_3	1.130042	0.264220	4.276900	0.0004
α_4	3558.117	1851.979	1.921252	0.0691
α_5	-3850.093	3498.033	-1.100645	0.2841
β_0	2349.653	3620.802	0.648932	0.5238
β_1	0.004260	0.681070	0.006255	0.9951
β_2	-2.002327	1.211622	-1.652600	0.1140
β_3	1.172795	0.275988	4.249440	0.0004
β_4	3511.283	1934.465	1.815119	0.0845
β_5	-4327.418	3653.833	-1.184350	0.2502

Equation:
$DEXP = \alpha_0 + \alpha_1 FDIIN + \alpha_2 FDIOUT + \alpha_3 DGDP + \alpha_4 DER + \alpha_5 D$
R-squared = 0.636776

Equation:
$DIMP = \beta_0 + \beta_1 FDIIN + \beta_2 FDIOUT + \beta_3 DGDP + \beta_4 DER + \beta_5 D$
R-squared = 0.642328

Table 2 Estimation results of Model 1

Model 2 was supposed to represent a simultaneous system of equations with three crucial variables mentioned in the first article DGDP, DEXP and FDIIN being endogenous. The purpose is not to describe the whole economy but to find up only if the variables in question have any quantitative linkage. Including exogenous variables DER, FDIOUT and D the exact identification of each equation in Model 2 is ensured and contemporaneously an economic interpretation of their parameters corresponds with the topic.

Model 2

$$DGDP = \alpha_0 + \alpha_1 DEXP + \alpha_2 FDIIN + \alpha_3 DER + \alpha_4 D + \alpha_5 FDIOUT + u$$

$$DEXP = \beta_0 + \beta_1 DGDP + \beta_2 FDIIN + \beta_3 DER + \beta_4 D + \beta_5 FDIOUT + v$$

$$FDIIN = \gamma_0 + \gamma_1 DGDP + \gamma_2 DEXP + \gamma_3 DER + \gamma_4 D + \gamma_5 FDIOUT + w$$

Unfortunately, the application of the 2SLS, resp. 3SLS, methods has failed due to a singular transformation matrix. That is why also the Model 2 was estimated with the help of SUR. Of course, the endo-/exogenous classification becomes irrelevant now.

The SUR method is used to gain efficiency when the equations are related through the error term. The estimation is performed in two steps. First step: each equation is estimated by OLS, and the residuals from the equations are used to estimate Σ. Second step: substitute Σ found in step 1 for Σ of GLS estimator. Detailed description e.g. in [3]. Using this method, the model is still treated as a system endowed with the same stochastic environment. The results of the SUR method application to Model 2 are in Table 3. The overall characteristics is similar as that of Model 1, acceptable R-squared but some of proposed explaining variables are inefficient. It means that the model does not confirm their influence on relevant endogenous variable.

Tables 2 and 3 are the authentic Eviews outputs in which the parameters are rewritten to correspond with Models 1 and 2. While the Prob. indicator of the t-statistics show directly a possible influence of relevant exogenous variable, the substitute versus complement question related to exports and FDI outcomes is answered by exploiting a sign of relevant coefficient. While substitutive relation takes place when increase in FDI is accompanied by decrease in export, and vice versa, complementarity means a movement in the same direction. In our context, parameter $\propto_2$ (at 10 % significance level) in Model 1, respective β_5 (5 %) in Model 2, have a negative sign what means that the entities are substitutes.

System: MODEL 2
Estimation Method: Seemingly Unrelated Regression

	Coefficient	Std. Error	t-Statistic	Prob.
α_0	-392.9648	2252.627	-0.174447	0.8627
α_1	0.613797	0.089127	6.886794	0.0000
α_2	-0.264915	0.417429	-0.634635	0.5305
α_3	-4085.859	748.0463	-5.462041	0.0000
α_4	4345.508	2051.065	2.118660	0.0425
α_5	1.524980	0.733490	2.079074	0.0463
β_0	2117.219	3428.144	0.617599	0.5415
β_1	1.469423	0.213368	6.886788	0.0000
β_2	0.170104	0.650322	0.261568	0.7954
β_3	5517.621	1628.140	3.388911	0.0020
β_4	-5843.903	3382.395	-1.727741	0.0943
β_5	-2.397366	1.148987	-2.086504	0.0455
β_5	4487.411	768.7011	5.837654	0.0000
γ_1	-0.093218	0.146884	-0.634636	0.5305
γ_2	0.025001	0.095587	0.261554	0.7955
γ_3	-1529.339	706.0051	-2.166186	0.0384
γ_4	1939.646	1316.268	1.473595	0.1510
γ_5	-0.230165	0.480742	-0.478770	0.6356

Equation:
$DGDP = \alpha_0 + \alpha_1 DEXP + \alpha_2 FDIIN + \alpha_3 DER + \alpha_4 D + \alpha_5 FDIOUT$
R-squared = 0.871218

Equation:
$DEXP = \beta_0 + \beta_1 DGDP + \beta_2 FDIIN + \beta_3 DER + \beta_4 D + \beta_5 FDIOUT$
R-squared = 0.599259

Equation:
$FDIIN = \gamma_0 + \gamma_1 DGDP + \gamma_2 DEXP + \gamma_3 DER + \gamma_4 D + \gamma_5 FDIOUT$
R-squared = 0.439335

Table 3 Estimation results of Model 2

3 Empirical Findings and Conclusions

The importance of a positive development of GDP, the exports and incoming investments for the Czech economic health is evident. To support their growth, CNB determines convenient macroeconomic quantities according to its own choice; after exhausting the impact of an inflation targeting, during 2014 – 2017 period the exchange rate rule was applied. Taking this 3 years regulation into account, two models were formulated to study relevant relations and influences.

The finally choice of SUR method is explained and its positive features are stressed. The results of applying the SUR method are summarized in Table 2 (Model 1) and Table 3 (Model2). The relevance of the outputs is conforming to some objective facts: The available annual data of Czech Republic hardly cover from 20 to 22 years including different turbulences as the continuation of transition period in late nineties and the 2007 – 09 financial crisis.

Estimated parameters show that the economic growth represented by GDP increments is supported by exports and exchange rate and also the regulation of the CZK/EUR parity evinces a positive effect. The incoming FDI do not affect significantly neither GDP nor exports. The amount of incoming FDI, as well as the export, evidently depends on the exchange rate but there is no apparent influence of its temporary regulation. For including out-coming FDI as an exogenous variable it also was found that FDI outcomes are substitutes of the exports in the Czech economic environment. At the same time, there is no relation between FDIIN, resp. FDIOUT, to the imports.

The period of exchange rate regulation is distinguished and its positive contribution to actual GDP growth cannot be denied. Nevertheless, the opinions exist, e.g. [8] that the overheating of the Czech economy in a near future will follow as a necessary consequence.

Acknowledgements

Financial support of IGA VŠE F4/78/2018 is gratefully acknowledged.

References

[1] Blanchard, O., Dell´Ariccia, G. & Mauro, P. (2010). *Rethinking Macroeconomic Policy*, International Monetary Fund, https://www.imf.org/external/pubs/ft/spn/2010/spn1003.pdf
[2] Chakrabarti, R. & Scholnick, B. (2002). Exchange Rate Regimes and Foreign Direct Investment Flows, Review of World Economics, 138, No. 1, pp 1 – 21
[3] Greene, W. H. (2003). *Econometric Analysis*, Prentice Hall
[4] Mariel, P. & Pánková, V. (2010). Exchange Rate Effects on Foreign Direct Investment Focusing on Central European Economies, *Journal of Economics*, 58, No. 2, pp 176 - 187
[5] Mishkin, F. S. (2011). Monetary Policy Strategy: Lessons From the Crisis, *NBER Working Paper* 16755, http://www.nber.org/papers/w16755
[6] Pánková, V. (2010) FDI and Export: Substitute or Complement? In: *Mathematical Methods in Economics 2010*. České Budějovice, University of South Bohemia, pp. 495–498.
[7] Rodrik, D. (2008). " The Real Exchange Rate and Economic Growth, http://j.mp/2oVwtIO
[8] Šichtařová, M. (2018). EKONOMIKA: Pro českou ekonomiku už není lék http://neviditelnypes.lidovky.cz/ekonomika-pro-ceskou-ekonomiku-uz-neni-lek-f2i-/p_ekonomika.aspx?c=A180317_204613_p_ekonomika_wag
[9] Zarotidias, G. & Mylonidis, N. (2005). FDI and Trade in the UK, http://www.etsg.org/ETSG2005/papers/Zarotiadis&Mylonidis.pdf

Associations among using special education programs in organizations: Application of loglinear models in HRM

Vladimír Přibyl[1], Lenka Komárková[2] and Pavel Pudil[3]

Abstract. Vocational education and training certainly belong among the determinants of success and competitiveness of organizations. The rapid technological development results inevitably in the fact that knowledge and skills of employees become obsolete and organizations are forced to invest not negligible financial means into various educational and training programs. The general aim of the study is to identify the associations between usage of various types of special educational programs and the fact whether employee considers such training programs as a benefit. The study was based on data set of 529 organizations collected via questionnaire in 2017. The association structure among the educational program types was examined by using loglinear models for multivariate contingency tables. We have found that the existence of special educational programs for elderly employees is conditionally independent from existence of the other forms of education programs as well as from the fact whether employee considers the training programs as a benefit. The found association structure further shows significant homogeneous conditional associations between some pairs of variables.

Keywords: further education, loglinear models, multivariate contingency tables, association structure.

JEL Classification: C21, M53
AMS Classification: 62H05, 62H17

1 Introduction

Professional development, particularly further education and training certainly belong among the determinants of success and competitiveness of organizations. Furthermore, training and further education are crucial HRM activities with a wide-range of implications, especially in the development of a talent pool of human capital. Organizations are investing not negligible financial means into various educational and training programs. The rapid technological development we are witnessing in the last decade results inevitably in the fact that knowledge and skills of older employees become obsolete as discussed by several authors, e.g. Beck [2], Lu, L. [9], Lallemand & Rycx [7]. This fact together with the ageing general population and workforce motivate companies and organizations to focus their educational and training activities also on special programs for older workers, or 50+ (generally considered as aged 55-64). In this context, we investigate whether the special programs of further education for the category 50+ are associated with other types of special education or if they have a particular position in the portfolio of all the educational programs.

Beside this specific goal, the general aim of the study is to identify the links between a usage of various types of special educational programs in organizations. Particularly, we want to identify associations between the existence of an educational program for managers, for talent development, for women, for elderly 50+ people, for working in an international environment and the fact whether employee considers such training programs as a benefit. The identification of these associations is important especially from the viewpoint of our broader project which concerns of identifying the key factors of organizations success related to further professional education as explained in the Conclusions.

The method we used here to identify the associations is based on the loglinear models. They were used for analysing associations in categorical data by several authors. DeLone [6] investigated the factors that affect the successful use of computer-based information systems (CBIS) by small business. Bhattacherjee et al. [4] assessed the associations of job and some individual factors with occupational injuries among employed people from a general population in north-eastern France. Vuong, Napier & Tran [20] attempted to explore relationships between three different dimensions of a business operation, namely the stages of business development, the methods of creativity and the major cultural values. A log-linear modelling for 3-dimensional contingency tables was used by Moreira [15] with categorical time series of standardized precipitation index for prediction of monthly drought

[1,2,3] University of Economics in Prague, Faculty of Management, Department of Exact Methods,
Jarošovská 1117/II, 377 01 Jindřichův Hradec, Czech Republic,
email: [1]pribyl@fm.vse.cz, [2]komarkol@fm.vse.cz, [3]pudil@fm.vse.cz

severity. Similar application to assessing homogeneous regions relative to drought was published by Li, Zhou & Hu [8]. Onder, M., Onder, S., & Mutlu [16] implemented the hierarchical loglinear analysis method to occupational fatalities occurred in the period of 1980-2004 in the five underground coal mines in Turkey. Bilder & Loughin [5] try to provide statistical analysis methods that can be applied to "choose all that apply" questions in complex survey sampling situations and show that loglinear models are developed to incorporate the multiple responses inherent in these types of questions. They applied this approach to data from the National Health and Nutrition Examination Survey. To conclude, though the applications of loglinear models for analyzing categorical data are not rare, to our knowledge they are absent in the area of education. The only paper from all those mentioned here, which is at least loosely related to education is that of DeLone [6].

2 Data and Methods

The study was based on the questionnaire distributed among the distance students of our faculty who often take positions of middle and senior management in their organizations. Data collection took place in the second half of 2017. The analysis was based on data set of 529 organizations operating in the Czech Republic and all six considered variables were in a dichotomous form (0-no/1-yes). Empirical distribution of such variables expressed by relative frequencies is presented in Figure 1, where the types of special education programs are ordered according to frequencies of their using. We can see that special education and training programs for managers are the most frequent (65.6%) and special education programs for elderly 50+ employees are the least frequent (1.1%). In addition, employees from 52.9% organizations in the sample have an education as a benefit. That means the second place.

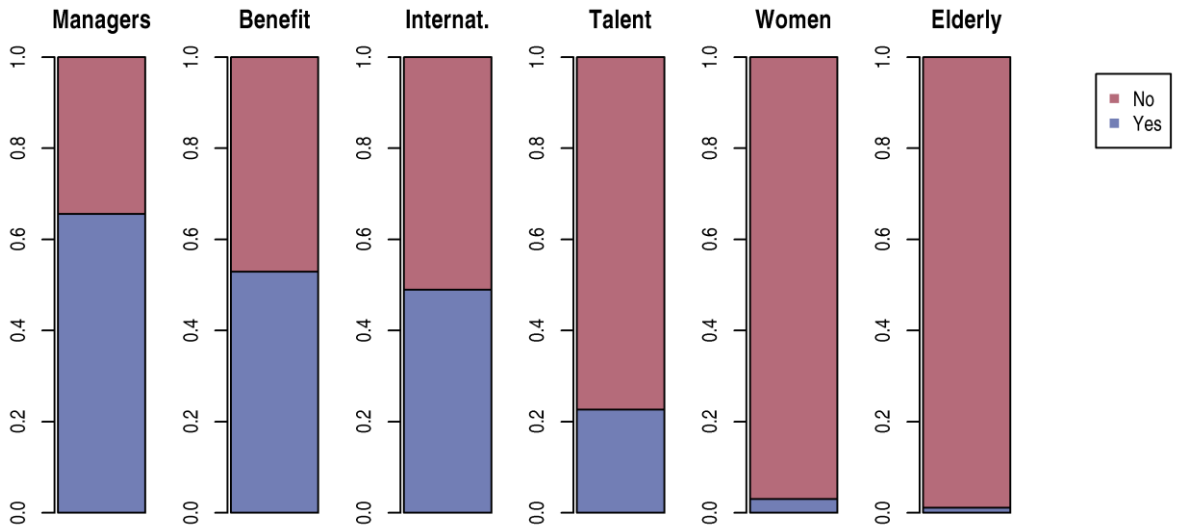


Figure 1 Empirical distribution of special education programs existence

The association structure among the variables was expressed by mosaic plots (Meyer et al. [11], Zeileis et al. [21]) and examined by using loglinear models for multivariate contingency tables (e.g., Agresti [1]) with cell frequencies as response and considered six factors as covariates. Statistical analysis was performed using the statistical software R (R Core Team [19]), R package vcd (Meyer et al. [10]). With respect to the sample size and the model's interpretation capability, the loglinear model with all three-way interactions was taken as the initial model. By the means of the likelihood ratio tests, insignificant terms were eliminated to arrive at the final model which describes the association structure among the six variables. Deviances of the final and the initial model differ by 21.79 on 29 degrees of freedom (p-value of the submodel test: $p = 0.829$).

3 Results and Discussion

Marginal associations were explored by matrix of mosaic plots for two-way contingency tables, see Figure 2**.** Analogously, for visualizing the six-way contingency table, mosaic plot in Figure 3 was used. Both plots also indicate that there is some association among the investigated variables (large absolute values of Pearson residuals from the independence model are indicated by colors – shades of blue and red).

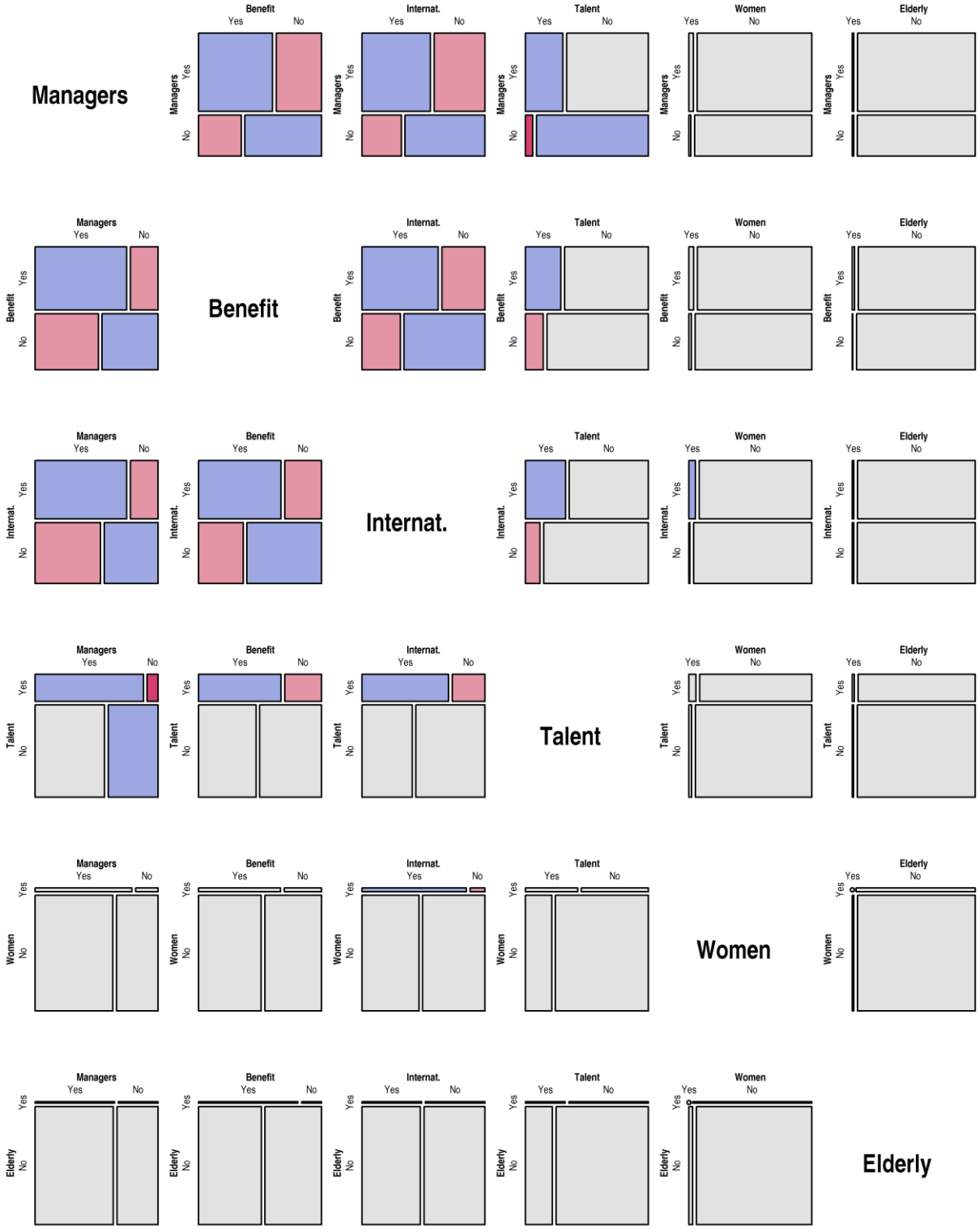


Figure 2 Matrix of mosaic plots

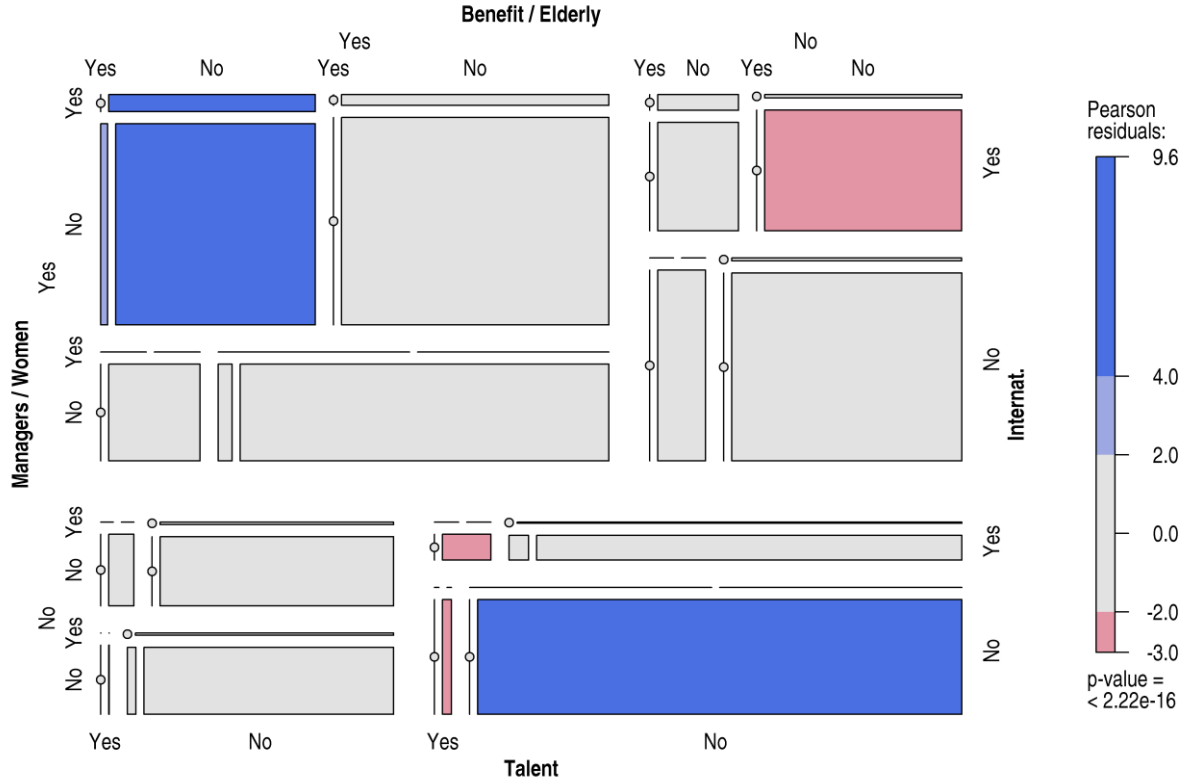


Figure 3 Mosaic plot under the independence model

No three-way interactions were present in the final model which involves six two-way interactions, see the left side of Table 1. Particularly, Table 1 reports estimated odds ratios (OR), related 95% Wald confidence intervals and p-values based on the likelihood-ratio tests. For each pair of variables from Table 1, their conditional (given the remaining variables) association structure is the same for all levels of the conditioning variables. That is, the pairs of variables from Table 1 express the homogeneous and significant associations with respect to the remaining variables. Those pairs not mentioned in Table 1 do not express significant conditional association. The results are further summarized in Table 2 which displays the revealed association structure in a matrix form. It clearly shows that usage of special educational programs for elderly employees has a special position. Their existence in an organization is conditionally independent from existence of the remaining forms of education programs as well as from the fact whether employee considers the training programs as a benefit.

	Conditional Association			Marginal Association		
Term	**OR**	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**
Managers*International	1.69	(1.12, 2.55)	0.012	2.73	(1.88, 3.98)	< 0.001
Managers*Talent	5.85	(3.02, 11.32)	< 0.001	7.12	(3.72, 13.64)	< 0.001
Managers*Benefit	2.37	(1.60, 3.50)	< 0.001	2.93	(2.02, 4.26)	< 0.001
Women*International	7.66	(1.72, 34.03)	< 0.001	7.66	(1.73, 34.03)	< 0.001
International*Talent	2.89	(1.82, 4.59)	< 0.001	3.63	(2.33, 5.68)	< 0.001
International*Benefit	3.24	(2.24, 4.70)	< 0.001	3.74	(2.61, 5.37)	< 0.001

Table 1 Conditional association based on the final loglinear model (left side) and marginal association for each pair of factors (right side). Note: The symbol * denotes an interaction.

Further, existence of educational programs for managers is associated with existence of educational programs for talent development and for working in an international environment. The strongest (and positive) conditional association (OR of 7.66) was found between existence of educational programs for women and for working in an international environment. Working in an international environment is further associated with existence of educational programs for talent development. Finally, the fact whether employee considers the educational programs as a benefit is significantly associated with existence of educational programs for managers and for working in an international environment. In addition, the estimated conditional associations are weaker than

marginal associations except one case (of the strongest association mentioned above) where the estimates for odds ratio are the same for both situations.

Type of Special Program	Managers	Women	Talent	Elderly	International	Benefit
Managers	x	No	Yes	No	Yes	Yes
Women	No	x	No	No	Yes	No
Talent	Yes	No	x	No	Yes	No
Elderly	No	No	No	x	No	No
International	Yes	Yes	Yes	No	x	Yes
Benefit	Yes	No	No	No	Yes	x

Table 2 Association structure of investigated variables

4 Conclusion

The results of this study demonstrate that as far as using various special education programs is concerned, the associations among them exist, in some cases quite strong. This fact has to be considered having in mind the overall objective of our broader research project "Organizations success related to professional education: identifying key factors". As the project title suggests, its ultimate goal is to identify among a number of variables or features characterizing the overall process of professional education just those variables which have a significant impact on organizations success. Then, such variables could be looked upon as the key factors of the economic success. The number of all the possible characteristics of the overall process of further professional education in organizations can be really a big one. We plan to use several alternative approaches to pursue the work on the project, namely the tests of independence, multiple regression analysis or logistic and ordinal regression, that have been already successfully used in our research (Pudil et al. [18], Mikova et al. [13], Mikova, Komarkova and Pudil [12], Mikova, Pudil and Komarkova [14]). Furthermore, in the following stages of the project we plan to use also the learning methods of feature selection and classification from statistical pattern recognition that helped to yield valuable results in the project concerned with identifying key factors of competitiveness of Czech companies (e.g. Pudil et al. [17]).

All these named approaches need some reduction of the number of variables entering into the analysis. It is true particularly for learning approaches from pattern recognition where the so called curse of dimensionality due to a famous American mathematician Richard Bellman (see Bellman [3]) is valid. The consequence is that to match the original dimensionality (the original number of characteristics or variables), we would need an enormous size of the data set, however its size is in most real problems rather limited. It means we need to reduce the number of variables characterizing the process of further professional education, while not losing the essential information. One of the ways to achieve this can be just by means of analysing the association structure of investigated variables (the usage of special education programs). Those variables having strong associations need not be considered as independent variables but can be grouped together.

Though, with respect to a low frequency of special programs for women and the category 50+ in our sample, the results of this pilot have to be taken with some caution, it is obvious that analyzing the associations among using special educations programs is a suitable method for our project. As we shall soon begin to analyze a much bigger data set concerning the profession education in our country, we can also expect higher numbers of using special education programs. Beside applying loglinear models for this purpose which was the aim of the current study we plan to use another approach based on data mining, namely the association rules. The applications of this approach are really numerous, they can be found even in the area of education or training.

Acknowledgements

This contribution is supported by the Czech Science Foundation (GA ČR) under the grant number 18-01159S.

References

[1] Agresti, A. (2003). *Categorical Data Analysis*. John Wiley & Sons Inc.

[2] Beck, V. (2013). Employers' use of older workers in the recession. *Employee Relations*, *35*(3), 257-271.

[3] Bellman, R,.E., (1957). Dynamic programming. Princeton University Press. ISBN 978-0-691-07951-6., Republished: Richard Ernest Bellman (2003). Dynamic Programming. Courier Dover Publications. ISBN 978-0-486-42809-3.

[4] Bhattacherjee, A., Chau, N., Sierra, C. O., Legras, B., Benamghar, L., Michaely, J. P., ... & Mur, J. M. (2003). Relationships of job and some individual characteristics to occupational injuries in employed people: a community-based study. *Journal of occupational health*, *45*(6), 382-391.

[5] Bilder, C. R., & Loughin, T. M. (2009). Modeling multiple-response categorical data from complex surveys. *Canadian Journal of Statistics*, *37*(4), 553-570.

[6] DeLone, W. H. (1988). Determinants of success for computer usage in small business. Mis Quarterly, 51-61

[7] Lallemand, T., & Rycx, F. (2009). Are older workers harmful for firm productivity?. *De Economist*, *157*(3), 273-292.

[8] Li, J., Zhou, S., & Hu, R. (2016). Hydrological drought class transition using SPI and SRI time series by loglinear regression. *Water resources management*, *30*(2), 669-684.

[9] Lu, L. (2012). Attitudes towards aging and older people's intentions to continue working: a Taiwanese study. *Career Development International*, *17*(1), 83-98.

[10] Meyer, D., Zeileis, A., Hornik, K. (2006). The Strucplot Framework: Visualizing Multi-Way Contingency Tables with vcd. *Journal of Statistical Software*, 17(3), 1-48.

[11] Meyer, D., Zeileis, A., Hornik, K. (2017). *vcd: Visualizing Categorical Data*. R package version 1.4-4.

[12] Míková, I., Komárková, L., & Pudil, P. (2017). Relation between Usage of More Effective Methods of Lifelong Learning and Evaluating Its Impacts. In L. G. Chova, A. L. Martínez, & I. C. Torres (Eds.), *Conference Proceedings: 11th International Technology, Education and Development Conference (INTED), Valencia, 6.-8.3.2017* (pp. 963–970). Valencia: International Academy of Technology, Education and Development.

[13] Mikova, I., Komarkova, L., Pudil, P., Pribyl, V. (2018). Effectiveness of Professional Development and Training with Respect to the Scope of Company Activities and Ownership. *Proceedings of International Conference on Management, Leadership and Governance, May 2018, Bangkok (in print)*

[14] Míková, I., Pudil, P., & Komárková, L. (2016). Adopting Proper Methods for Studying Management in Post-Transformation Economies and Their Effect on the Employment Rate. *Turkish Online Journal of Educational Technology*, *Special Issue*, 218–226.

[15] Moreira, E. E. (2016). SPI drought class prediction using log-linear models applied to wet and dry seasons. *Physics and Chemistry of the Earth, Parts A/B/C*, *94*, 136-145.

[16] Onder, M., Onder, S., & Mutlu, A. (2012). Determination of noise induced hearing loss in mining: an application of hierarchical loglinear modelling. *Environmental monitoring and assessment*, *184*(4), 2443-2451.

[17] Pudil, P., Blažek, L., Částek, O., Somol, P., Pokorná, J., Králová, M. (2014). Identifying Corporate Performance Factors Based on Feature Selection in Statistical Pattern Recognition: METHODS, APPLICATION, INTERPRETATION. Brno: Munipress.

[18] Pudil, P., Dvorak, J., Komarkova, L., Mikova, I., Pribyl, V. (2018). Analysis of Investment into Further Education as a Part of International and Domestic HRM. *Proceedings of International Conference on Management, Leadership and Governance, May 2018, Bangkok (in print)*

[19] R Core Team (2018). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

[20] Vuong, Q. H., Napier, N. K., & Tran, T. D. (2013). A categorical data analysis on relationships between culture, creativity and business stage: the case of Vietnam. *International Journal of Transitions and Innovation Systems*, *3*(1), 4-24.

[21] Zeileis, A., Meyer, D., Hornik, K. (2007). Residual-based Shadings for Visualizing (Conditional) Independence. *Journal of Computational and Graphical Statistics*, 16(3), 507-525.

Fuzzy Decision Matrix in Case of Two Risk Factors

Ondřej Pavlačka[1], Vojtěch Sukač[2]

Abstract. A decision matrix models a particular problem of decision-making under risk. Elements of the matrix express outcomes of alternatives if the particular states of the world occur. The alternatives are usually compared on the basis of the expected values and variances of their outcomes. In the paper, we study the case where the states of the world are given by possible combinations of the vaguely defined states of two risk factors. We assume that probability distributions of the risk factors are known and their states are expressed by fuzzy sets. Furthermore, we consider the outcomes of alternatives under the particular states of the world to be expressed by fuzzy numbers. We show that in such a case a fuzzy decision matrix can be understood as a collection of fuzzy rule-based systems. We derive the proper formulas for computing the fuzzy expected values and variances of the outcomes of alternatives.

Keywords: Decision matrices, decision-making under risk, fuzzy states of the world, fuzzy rule-based systems.

JEL classification: C44
AMS classification: 90B50

1 Introduction

In decision making under risk, a decision matrix (see Tab. 1) is a popular tool of risk analysis (see e.g. [1, 13]). It describes how outcomes of alternatives $x_1, \ldots, x_n$ depend on the fact which of the 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 are given by $p_1, \ldots, p_m$. Thus, for any $i \in \{1, \ldots, n\}$, the outcome of choosing an alternative x_i is expressed by a discrete random variable H_i that takes on the values $h_{i,1}, \ldots, h_{i,m}$ with the 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 $var\ H_1, \ldots, var\ H_n$ of their outcomes, where $EH_i = \sum_{j=1}^{m} p_j \cdot h_{i,j}$ and $var\ H_i = \sum_{j=1}^{m} p_j \cdot (h_{i,j} - EH_i)^2$, $i = 1, \ldots, m$.

	S_1	S_2	$\cdots$	S_m		
	p_1	p_2	$\cdots$	p_m		
x_1	$h_{1,1}$	$h_{1,2}$	$\cdots$	$h_{1,m}$	EH_1	$var\ H_1$
x_2	$h_{2,1}$	$h_{2,2}$	$\cdots$	$h_{2,m}$	EH_2	$var\ H_2$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_n	$h_{n,1}$	$h_{n,2}$	$\cdots$	$h_{n,m}$	EH_n	$var\ H_n$

Table 1 Decision matrix.

In practical applications, the states of the world as well as the outcomes of alternatives can be determined vaguely (see e.g. [2] and references therein). The states of the world are mostly described verbally, like "moderate increase of the gross domestic product", "low inflation rate", etc. As for the outcomes of alternatives, it can be sometimes problematic to express them precisely because we may not have enough information. For instance, the evaluation under a certain state of the world can be described as "increase about 2 %". In some cases, it is more natural for a decision-maker to express the evaluations by selecting a term from a given linguistic scale.

The vaguely described pieces of information can be mathematically modelled by means of tools of fuzzy sets theory. A decision matrix with fuzzy states of the world and fuzzy outcomes of alternatives is called a fuzzy decision matrix. Talašová and Pavlačka [12] introduced a model where the fuzzy states of the world are expressed 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 the crisp (i.e. exactly described) states of the world; they set the probabilities of the fuzzy states of the world applying the formula proposed by Zadeh in [15]. Within this approach, the outcomes of alternatives are understood as discrete random variables taking on fuzzy values with the probabilities of the fuzzy states of the world. In [4, 7], it is shown that the Zadeh's probabilities of fuzzy events lack the common

[1]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
[2]dtto, vojta.sukac@gmail.com

interpretation of a probability measure. Therefore, an alternative to how the information contained in a fuzzy decision matrix can be treated was proposed by Rotterová and Pavlačka in [5, 8]. The way is based on the idea that a fuzzy decision matrix does not determine discrete fuzzy random variables, but a collection of fuzzy rule-based systems (a fuzzy rule-based system was introduced by Zadeh in [16]).

In real world, there are situations where the outcomes of alternatives are affected by several risk factors. In the paper, we will study the case where the states of the world are given by possible combinations of the vaguely defined states of two risk factors. We assume that probability distributions of the risk factors are known and their states are expressed by fuzzy sets. The case where the two risk factors are continuous was already considered by Talašová and Pavlačka in [12]. However, as it is mentioned above, they proposed to apply Zadeh's probabilities of fuzzy events. In concordance with recent developments, we will construct the corresponding collection of fuzzy rule-based systems and derive the proper formulas for computing the fuzzy expected values and variances of the outcomes of alternatives.

The rest of the paper is organised as follows. The way how can be vaguely defined states of the world and uncertain outcomes of alternatives expressed by means of tools of fuzzy sets theory is presented in next section. In Section 3, the case where the states of the world are given by combinations of the states of two risk factors is analysed.

2 Fuzzy states of the world and fuzzy outcomes of alternatives

First, let us describe how the vaguely defined states of the world and uncertain outcomes of alternatives can be expressed by means of tools of fuzzy sets theory.

A *fuzzy set* A on a non-empty universal set U is determined by its membership function $\mu_A : U \to [0,1]$. Let us denote the family of all fuzzy sets on U by $\mathcal{F}(U)$. For any $\alpha \in (0,1]$, A_α denotes an α-cut of A, i.e. $A_\alpha := \{u \in U \mid \mu_A(u) \geq \alpha\}$. Let us note that any crisp set $A \subseteq U$ can be viewed as a fuzzy set of a special kind; the membership function μ_A coincides in such a case with the characteristic function χ_A of A, which implies that $A_\alpha = A$ for all $\alpha \in (0,1]$. This convention further allows us to consider crisp states of the world as a special kind of fuzzy states of the world.

Let us assume that a probability space $(\Omega, \mathcal{A}, P)$ is given, where Ω denotes a nonempty set of all elementary events (all possible states of the world in the future), $\mathcal{A}$ 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 $P(A) \in [0,1]$. Vaguely defined states of the world can be appropriately expressed (see [5, 8, 12]) by fuzzy sets $S_1, \ldots, S_m$ on Ω whose membership functions $\mu_{S_1}, \ldots, \mu_{S_m}$ are $\mathcal{A}$-measurable, i.e. the α-cuts $S_{j\alpha} \in \mathcal{A}$ for any $\alpha \in (0,1]$, $j = 1, \ldots, m$, and that form a fuzzy partition of Ω, i.e. $\sum_{j=1}^{m} \mu_{S_j}(\omega) = 1$ for any $\omega \in \Omega$. Such fuzzy sets $S_1, \ldots, S_m$ usually represent mathematical meanings of linguistic terms from a given linguistic scale (see [8]).

As was mentioned in Introduction, it can be sometimes difficult to determine outcomes of alternatives under particular states of the world by real numbers. As it is stated in [8], one reason can be the lack of information caused e.g. by inaccuracies of measurements or a lower quality of data transmissions. Another reason can be that it is more natural for a decision-maker to describe the outcomes linguistically rather than by numbers.

Uncertain quantities can be modelled by fuzzy numbers. A *fuzzy number* X is a fuzzy set on the set of all real numbers $\mathbb{R}$ whose membership function $\mu_X : \mathbb{R} \to [0,1]$ fulfils the following three conditions: there exists at least one $x \in \mathbb{R}$ such that $\mu_X(x) = 1$, the α-cuts X_α are closed intervals for any $\alpha \in (0,1]$, and the set $Supp\ X := \{x \in \mathbb{R} \mid \mu_X(x) > 0\}$, called the support of X, is bounded. The set of all fuzzy numbers will be denoted by $\mathcal{F}_N(\mathbb{R})$.

In [3], it was shown that any fuzzy number X can be uniquely given by the pair of functions $\underline{x}$ and $\overline{x}$ defined on $[0,1]$ such that $[\underline{x}(\alpha), \overline{x}(\alpha)] = X_\alpha$ for all $\alpha \in (0,1]$ and $[\underline{x}(0), \overline{x}(0)]$ means the closure of the support of X. For such a description of a fuzzy number, the notation $X = (\underline{x}, \overline{x})$ will be used throughout the paper. A real number $x \in \mathbb{R}$ can be viewed as a fuzzy number $X = (\underline{x}, \overline{x})$, where $\underline{x}(\alpha) = \overline{x}(\alpha) = x$ for all $\alpha \in [0,1]$. By this convention we can handle the cases where some outcomes of alternatives are crisp and some are fuzzy.

There are two ways of determining fuzzy outcomes of alternatives $H_{i,j}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. The first way is to specify them directly by fuzzy numbers. For instance, some expert can described the outcome of an alternative directly by a fuzzy number expressing the uncertain value "about 5 % profit" . The second possibility of expressing the fuzzy evaluation of the alternative consists in the fact that the outcome is represented by a linguistic variable (linguistic variables were introduced in [16]). A decision-maker evaluates the outcomes of alternatives under the particular states of the world by appropriate linguistic terms whose mathematical meanings are described by fuzzy numbers.

3 Fuzzy decision matrix in case of two risk factors

In this section, let us study the case where outcomes of alternatives are affected by two risk factors denoted by R_1 and R_2. First, we construct the corresponding collection of fuzzy rule-based systems, and then derive the proper formulas for computing the fuzzy expected values and variances of the outcomes of alternatives.

Let us assume that probability spaces $(\Omega_1, \mathcal{A}_1, P_1)$ and $(\Omega_2, \mathcal{A}_2, P_2)$ are given, where for $j = 1, 2$, Ω_j denotes a nonempty set of all possible states of the j-th risk factor, $\mathcal{A}_j$ forms a σ-algebra of random events on Ω_j, and $P_j : \mathcal{A}_j \to [0, 1]$ is a probability measure. Vaguely defined states of the j-th risk factor are expressed by fuzzy sets $S_{j1}, \ldots, S_{jm_j}$ on Ω_j, whose membership functions $\mu_{S_{j1}}, \ldots, \mu_{S_{jm_j}}$ are $\mathcal{A}_j$-measurable, and that form a fuzzy partition of Ω_j, i.e. $\sum_{k=1}^{m_j} \mu_{S_{jk}}(\omega_j) = 1$ for any $\omega_j \in \Omega_j$. The fuzzy states of the world are then represented by the Cartesian products $S_{1k} \times S_{2l}$, $k = 1, \ldots, m_1$, $l = 1, \ldots, m_2$, where

$$\mu_{S_{1k} \times S_{2l}}(\omega_1, \omega_2) = \min\{\mu_{S_{1k}}(\omega_1)), \mu_{S_{2l}}(\omega_2))\}, \quad \text{for any } (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2.$$

For $i = 1, \ldots, n$, the outcomes of an alternative x_i under the particular states of the worlds are expressed by fuzzy numbers $H_{i,(kl)} = (\underline{h}_{i,(kl)}, \overline{h}_{i,(kl)})$, $k = 1, \ldots, m_1$, $l = 1, \ldots, m_2$. Information about how the outcomes of alternatives depend on the two risk factors can be described by the fuzzy decision matrix presented in Tab. 2.

	$S_{11} \times S_{21}$	$S_{11} \times S_{22}$	$\cdots$	$S_{1m_1} \times S_{2m_2}$
x_1	$H_{1,(11)}$	$H_{1,(12)}$	$\cdots$	$H_{1,(m_1m_2)}$
x_2	$H_{2,(11)}$	$H_{2,(12)}$	$\cdots$	$H_{2,(m_1m_2)}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_n	$H_{n,(11)}$	$H_{n,(12)}$	$\cdots$	$H_{n,(m_1m_2)}$

Table 2 Fuzzy decision matrix in case of two risk factors.

In concordance with the procedure proposed in [5, 8], the information given in Tab. 2 about the outcome of an alternative x_i, $i \in \{1, \ldots, n\}$, can be transformed into the following basis of fuzzy If-Then rules:

$$\begin{aligned}
&\text{If the risk factor } R_1 \text{ is } S_{11} \text{ and the risk factor } R_2 \text{ is } S_{21}, \text{ then the outcome of } x_i \text{ is } H_{i,(11)}.\\
&\text{If the risk factor } R_1 \text{ is } S_{11} \text{ and the risk factor } R_2 \text{ is } S_{22}, \text{ then the outcome of } x_i \text{ is } H_{i,(12)}.\\
&\ldots\\
&\text{If the risk factor } R_1 \text{ is } S_{1m_1} \text{ and the risk factor } R_2 \text{ is } S_{2m_2}, \text{ then the outcome of } x_i \text{ is } H_{i,(m_1m_2)}.
\end{aligned} \tag{1}$$

For computing the output of the fuzzy rule-based system (1), it was shown in [8] that it should be applied the *generalised Sugeno's method of fuzzy inference* introduced by Talašová [11]. It represents the extension of the *Sugeno's method of fuzzy inference*, introduced by Sugeno [10], that on the right-hand side of the rules considers outcomes given by real numbers instead of fuzzy numbers. For any $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, the outcome of choosing x_i is given by a fuzzy number $H_i(\omega_1, \omega_2)$ that is obtained as a weighted average of fuzzy numbers $H_{i,(11)}, H_{i,(12)}, \ldots, H_{i,(m_1m_2)}$ as follows:

$$H_i(\omega_1, \omega_2) = \frac{\sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \min\{\mu_{S_{1,k}}(\omega_1), \mu_{S_{2,l}}(\omega_2)\} \cdot H_{i,(kl)}}{\sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \min\{\mu_{S_{1,k}}(\omega_1), \mu_{S_{2,l}}(\omega_2)\}}, \tag{2}$$

i.e. if we denote $H_i(\omega_1, \omega_2) = (\underline{h}_{i,\omega_1,\omega_2}, \overline{h}_{i,\omega_1,\omega_2})$, then for all $\alpha \in [0, 1]$:

$$\begin{aligned}
\underline{h}_{i,\omega_1,\omega_2}(\alpha) &= \frac{\sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \min\{\mu_{S_{1,k}}(\omega_1), \mu_{S_{2,l}}(\omega_2)\} \cdot \underline{h}_{i,(kl)}(\alpha)}{\sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \min\{\mu_{S_{1,k}}(\omega_1), \mu_{S_{2,l}}(\omega_2)\}},\\
\overline{h}_{i,\omega_1,\omega_2}(\alpha) &= \frac{\sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \min\{\mu_{S_{1,k}}(\omega_1), \mu_{S_{2,l}}(\omega_2)\} \cdot \overline{h}_{i,(kl)}(\alpha)}{\sum_{k=1}^{m_1} \sum_{l=1}^{m_2} \min\{\mu_{S_{1,k}}(\omega_1), \mu_{S_{2,l}}(\omega_2)\}}.
\end{aligned}$$

Thus, for $i = 1, \ldots, n$, the fuzzy outcome of an alternative x_i is expressed by the mapping $H_i : \Omega_1 \times \Omega_2 \to \mathcal{F}_N(\mathbb{R})$ given by (2). Since we operate within the given probability spaces $(\Omega_1, \mathcal{A}_1, P_1)$ and $(\Omega_2, \mathcal{A}_2, P_2)$, $H_1, \ldots, H_n$ represent fuzzy random variables. Analogously as in the case of a crisp decision matrix described by Tab. 1, the alternatives $x_1, \ldots, x_n$ can be compared on the basis of the fuzzy expected values and fuzzy variances of $H_1, \ldots, H_n$. Let us derive now the formulas for their computation.

Let P be the joint probability distribution of the risk factors R_1 and R_2. For the sake of simplicity in notation, let

$$s\left(\omega_1,\omega_2;h_{i,(11)},\ldots,h_{i,(m_1m_2)}\right) := \frac{\sum_{k=1}^{m_1}\sum_{l=1}^{m_2}\min\{\mu_{S_{1,k}}(\omega_1),\mu_{S_{2,l}}(\omega_2)\}\cdot h_{i,(kl)}}{\sum_{k=1}^{m_1}\sum_{l=1}^{m_2}\min\{\mu_{S_{1,k}}(\omega_1),\mu_{S_{2,l}}(\omega_2)\}}.$$

For $i=1,\ldots,n$, the fuzzy expected value of H_i is a fuzzy number $EH_i=\left(\underline{Eh}_i,\overline{Eh}_i\right)$ where for all $\alpha\in[0,1]$:

$$\begin{aligned}\underline{Eh}_i(\alpha) &= \min\left\{\int_{\Omega_1\times\Omega_2} s\left(\omega_1,\omega_2;h_{i,(11)},\ldots,h_{i,(m_1m_2)}\right)dP \mid h_{i,(k,l)}\in[\underline{h}_{i,(kl)}(\alpha),\overline{h}_{i,(kl)}(\alpha)],\right.\\ &\left. k=1,\ldots,m_1,\ l=1,\ldots,m_2\right\} = \int_{\Omega_1\times\Omega_2} s\left(\omega_1,\omega_2;\underline{h}_{i,(11)},\ldots,\underline{h}_{i,(m_1m_2)}\right)dP, \qquad (3)\\ \overline{Eh}_i(\alpha) &= \max\left\{\int_{\Omega_1\times\Omega_2} s\left(\omega_1,\omega_2;h_{i,(11)},\ldots,h_{i,(m_1m_2)}\right)dP \mid h_{i,(k,l)}\in[\underline{h}_{i,(kl)}(\alpha),\overline{h}_{i,(kl)}(\alpha)],\right.\\ &\left. k=1,\ldots,m_1,\ l=1,\ldots,m_2\right\} = \int_{\Omega_1\times\Omega_2} s\left(\omega_1,\omega_2;\overline{h}_{i,(11)},\ldots,\overline{h}_{i,(m_1m_2)}\right)dP. \qquad (4)\end{aligned}$$

The fuzzy variance of H_i is a fuzzy number $var\ H_i=\left(\underline{var\ h}_i,\overline{var\ h}_i\right)$ that can be obtained as follows: Let us denote

$$\begin{aligned}&v\left(h_{i,(11)},\ldots,h_{i,(m_1m_2)}\right)\\ &:=\int_{\Omega_1\times\Omega_2}\left(s\left(\omega_1,\omega_2;h_{i,(11)},\ldots,h_{i,(m_1m_2)}\right)-\int_{\Omega_1\times\Omega_2} s\left(\omega_1,\omega_2;h_{i,(11)},\ldots,h_{i,(m_1m_2)}\right)dP\right)^2 dP. \qquad (5)\end{aligned}$$

Then for all $\alpha\in[0,1]$:

$$\underline{var\ h}_i(\alpha)=\min\left\{v\left(h_{i,(11)},\ldots,h_{i,(m_1m_2)}\right)\mid h_{i,(k,l)}\in[\underline{h}_{i,(kl)}(\alpha),\overline{h}_{i,(kl)}(\alpha)],k=1,\ldots,m_1,l=1,\ldots,m_2\right\}, \qquad (6)$$

$$\overline{var\ h}_i(\alpha)=\max\left\{v\left(h_{i,(11)},\ldots,h_{i,(m_1m_2)}\right)\mid h_{i,(k,l)}\in[\underline{h}_{i,(kl)}(\alpha),\overline{h}_{i,(kl)}(\alpha)],k=1,\ldots,m_1,l=1,\ldots,m_2\right\}. \qquad (7)$$

Let us note that these optimization problems can be very difficult to solve. The algorithms for solving these problems could be a subject of further research.

Example 1. Let us assume that the risk factors R_1 and R_2 are continuous (their domain is $\mathbb{R}$) and mutually independent. Let their probability distributions be given by density functions f_1 and f_2, respectively. The formulas (3) and (4) for computing the fuzzy expected values $EH_i=\left(\underline{Eh}_i,\overline{Eh}_i\right)$, $i=1,\ldots,n$, are then expressed for all $\alpha\in[0,1]$ as follows:

$$\begin{aligned}\underline{Eh}_i(\alpha) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s\left(x_1,x_2;\underline{h}_{i,(11)},\ldots,\underline{h}_{i,(m_1m_2)}\right)f_1(x_1)f_2(x_2)dx_1dx_2,\\ \overline{Eh}_i(\alpha) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s\left(x_1,x_2;\overline{h}_{i,(11)},\ldots,\overline{h}_{i,(m_1m_2)}\right)f_1(x_1)f_2(x_2)dx_1dx_2.\end{aligned}$$

The auxiliary function v defined by (5), applied in the formulas (6) and (7) for computing the fuzzy variances $var\ H_1,\ldots,var\ H_n$, is in this case given by

$$v\left(\mathbf{h}_i\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(s\left(x_1,x_2;\mathbf{h}_i\right)-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s\left(y_1,y_2;\mathbf{h}_i\right)f_1(y_1)f_2(y_2)dy_1dy_2\right)^2 f_1(x_1)f_2(x_2)dx_1dx_2,$$

where $\mathbf{h}_i=\left(h_{i,(11)},\ldots,h_{i,(m_1m_2)}\right)$.

Remark 1. It can be easily seen from (2) that in the case where the states of the risk factors R_1 and R_2 are crisp, i.e. $S_{11},\ldots,S_{1m_1}\in\mathcal{A}_1$, and $S_{21},\ldots,S_{2m_2}\in\mathcal{A}_2$, respectively, and where the outcomes of alternatives are given by real numbers $h_{i,(kl)}$, $H_1,\ldots,H_n$ are discrete random variables taking on the values $h_{i,(11)},\ldots,h_{i,(m_1m_2)}$, $i=1,\ldots,n$, with the probabilities $p_{kl}=P(R_1=S_{1k},R_2=S_{2l})$, $k=1,\ldots,m_1$, $l=1,\ldots,m_2$. Hence, the approach presented in the paper represents an extension of a decision matrix.

Comparison of alternatives $x_1,\ldots,x_n$ can be based on the visual judgment of the membership functions of the fuzzy expected values $EH_1,\ldots,EH_n$ and of the fuzzy variances $var\ H_1,\ldots,var\ H_n$. Another possibility is to apply some defuzzification procedure, e.g. compare these fuzzy characteristics on the basis of their centres of gravity. We can also employ the linguistic approximation for characterization the fuzzy expected values and fuzzy variances, as was proposed by Talašová and Pavlačka [12].

Example 2. Let us consider the following situation: We can sell either only a product A (an alternative x_1), or two products A and B simultaneously (an alternative x_2). The degree of our satisfaction from this activity (derived not only from the future yield but also from the time we will spend by the activity) depend on two discrete independent risk factors – the demands for the products A and B. Let us assume that four levels of both demands are possible in the future, i.e. we have the couple of universes $\Omega_i = \{\omega_{i1}, \omega_{i2}, \omega_{i3}, \omega_{i4}\}$, $i = 1, 2$. The probabilities of the elementary events are given as follows: $P(\{\omega_{ij}\}) = p_{ij}$, $i = 1, 2$, $j = 1, \ldots, 4$, where $p_{11} = p_{14} = 0.15$, $p_{12} = p_{13} = 0.35$, and $p_{21} = p_{22} = p_{23} = p_{24} = 0.25$.

For both risk factors, we can distinguish two cases (fuzzy states of the risk factors) – "a small demand" (denoted by S_{i1}, $i = 1, 2$) and "a big demand" (denoted by S_{i2}, $i = 1, 2$). The corresponding fuzzy sets are given in the following way: $S_{i1} = \left\{{}^{1}|_{\omega_{i1}}, {}^{0.667}|_{\omega_{i2}}, {}^{0.333}|_{\omega_{i3}}, {}^{0}|_{\omega_{i4}}\right\}$, $S_{i2} = \left\{{}^{0}|_{\omega_{i1}}, {}^{0.333}|_{\omega_{i2}}, {}^{0.667}|_{\omega_{i3}}, {}^{1}|_{\omega_{i4}}\right\}$, $i = 1, 2$, where elements of the sets are in the form ${}^{\mu_{S_{ik}}(\omega_{ij})}|_{\omega_{ij}}$, $i, k = 1, 2$, $j = 1, \ldots, 4$.

How our degree of satisfaction, defined on the unit interval $[0, 1]$, depends on the fact which combination of demands for A and B will occur in the future is described in Tab. 3. The uncertain degrees of satisfaction $H_{1,(11)}, \ldots, H_{2,(22)}$ are for the sake of better clearness given by intervals (special kinds of fuzzy numbers whose α-cuts are the same).

	$S_{11} \times S_{21}$	$S_{11} \times S_{22}$	$S_{12} \times S_{21}$	$S_{12} \times S_{22}$
x_1	$[0.15, 0.35]$	$[0.15, 0.35]$	$[0.65, 0.85]$	$[0.65, 0.85]$
x_2	$[0, 0.15]$	$[0.4, 0.6]$	$[0.4, 0.6]$	$[0.85, 1]$

Table 3 Fuzzy decision matrix with interval-valued degrees of satisfactions $H_{1,(11)}, \ldots, H_{2,(22)}$.

Now, let us compute the expected interval degrees of satisfaction $EH_1 = [\underline{Eh}_1, \overline{Eh}_1]$ and $EH_2 = [\underline{Eh}_2, \overline{Eh}_2]$. Let us denote $H_{1,(11)} = [\underline{h}_{1,(11)}, \overline{h}_{1,(11)}], \ldots, H_{2,(22)} = [\underline{h}_{2,(22)}, \overline{h}_{2,(22)}]$, i.e. $\underline{h}_{1,(11)} = 0.15$, $\overline{h}_{1,(11)} = 0.35$, etc. Then according to (3) and (4), we get

$$\underline{Eh}_1 = \sum_{i=1}^{4}\sum_{j=1}^{4} p_{1i} \cdot p_{2j} \cdot \frac{\sum_{k=1}^{2}\sum_{l=1}^{2} \min\{\mu_{S_{1k}}(\omega_{1i}), \mu_{S_{2l}}(\omega_{2j})\} \cdot \underline{h}_{1,(kl)}}{\sum_{k=1}^{2}\sum_{l=1}^{2} \min\{\mu_{S_{1k}}(\omega_{1i}), \mu_{S_{2l}}(\omega_{2j})\}} = 0.4,$$

$$\overline{Eh}_1 = \sum_{i=1}^{4}\sum_{j=1}^{4} p_{1i} \cdot p_{2j} \cdot \frac{\sum_{k=1}^{2}\sum_{l=1}^{2} \min\{\mu_{S_{1k}}(\omega_{1i}), \mu_{S_{2l}}(\omega_{2j})\} \cdot \overline{h}_{1,(kl)}}{\sum_{k=1}^{2}\sum_{l=1}^{2} \min\{\mu_{S_{1k}}(\omega_{1i}), \mu_{S_{2l}}(\omega_{2j})\}} = 0.6,$$

and analogously, $EH_2 = [0.413, 0.588]$. We can see, that the expected degrees of satisfaction are very similar, the second one is less uncertain.

The interval variations $var\ H_1 = [\underline{var\ h}_1, \overline{var\ h}_1]$ and $var\ H_2 = [\underline{var\ h}_2, \overline{var\ h}_2]$ are calculated according to (6) and (7). By solving the following optimization problems for $r = 1, 2$:

$$\begin{aligned}
\underline{var\ h}_r &= \min\left\{ \sum_{i=1}^{4}\sum_{j=1}^{4} p_{1i} \cdot p_{2j} \cdot \frac{\sum_{k=1}^{2}\sum_{l=1}^{2} \min\{\mu_{S_{1k}}(\omega_{1i}), \mu_{S_{2l}}(\omega_{2j})\} \cdot h_{r,(kl)}}{\sum_{k=1}^{2}\sum_{l=1}^{2} \min\{\mu_{S_{1k}}(\omega_{1i}), \mu_{S_{2l}}(\omega_{2j})\}} \;\middle|\right. \\
&\qquad \left. h_{r,(k,l)} \in H_{r,(kl)},\ k = 1, 2,\ l = 1, 2 \right\}, \\
\overline{var\ h}_r &= \max\left\{ \sum_{i=1}^{4}\sum_{j=1}^{4} p_{1i} \cdot p_{2j} \cdot \frac{\sum_{k=1}^{2}\sum_{l=1}^{2} \min\{\mu_{S_{1k}}(\omega_{1i}), \mu_{S_{2l}}(\omega_{2j})\} \cdot h_{r,(kl)}}{\sum_{k=1}^{2}\sum_{l=1}^{2} \min\{\mu_{S_{1k}}(\omega_{1i}), \mu_{S_{2l}}(\omega_{2j})\}} \;\middle|\right. \\
&\qquad \left. h_{r,(k,l)} \in H_{r,(kl)},\ k = 1, 2,\ l = 1, 2 \right\},
\end{aligned}$$

we obtain $var\ H_1 = [0.008, 0.042]$ and $var\ H_2 = [0.026, 0.061]$. Thus, as it could be expected from Tab. 3, the variance of the degree of satisfaction by the second alternative x_2 is greater.

4 Conclusion

We have dealt with the problem of the extension of a decision matrix for the case where the states of the world are given by possible combinations of the vaguely defined states of two risk factors and the outcomes of alternatives

under the particular states of the world are expressed by fuzzy numbers. We have assumed that probability distributions of the two risk factors are known and the states of the risk factors are expressed by fuzzy sets. In concordance with recent developments in this field, we have constructed the corresponding collection of fuzzy rule-based systems (1) and derived the proper formulas for computing the fuzzy expected values and fuzzy variances of the outcomes of alternatives, based on which the alternatives can be compared.

Future work in this field will be focused on the case where the underlying probability measure is fuzzy. For instance, the parameters of the underlying probability distributions of the two risk factors, like μ and σ in the case of the normal distribution, could be given by fuzzy numbers (e.g. they can be estimated on the basis of fuzzy data, see Viertl [14]). Another possibility is the situation where the risk factors are of discrete nature and the probabilities of the particular states are fuzzy (e.g. expertly set). Such a problem in case of one risk factor was studied by Pavlačka and Rotterová in [6, 9].

Acknowledgements

Supported by the grant No. IGA_PrF_2018_024 *Mathematical Models* of the *Internal Grant Agency of Palacký University Olomouc*.

References

[1] Chankong, V. & Haimes, Y.Y. (1983). *Multiobjective decision making: theory and methodology*. New York: Elsevier Science B.V.

[2] Özkan, I. & Türken, I.B. (2014). Uncertainty and fuzzy decisions. In S. Banerjee, S. Sule Ercetin & A. Tekin (Eds.), *Chaos Theory in Politics* (pp. 17–27). Dordrecht: Springer Science+Business Media.

[3] Pavlačka, O. & Talašová, J. (2010). Fuzzy vectors as a tool for modeling uncertain muntidimensional quantities. *Fuzzy Sets and Systems*, 161 (11), 1585–1603.

[4] Pavlačka, O. & Rotterová, P. (2014). Probability of fuzzy events. In J. Talašová, J. Stoklasa & T. Talášek (Eds.), *Proceedings of the 32nd International Conference Mathematical Methods in Economics* (pp. 760–765). Olomouc: Palacký University Olomouc.

[5] Pavlačka, O. & Rotterová, P. (2016). Fuzzy decision matrices viewed as fuzzy rule-based systems. In A. Kocourek & M. Vavroušek (Eds.), *Proceedings of the 34th International Conference Mathematical Methods in Economics* (pp. 641–646). Liberec: Technical University of Liberec.

[6] Pavlačka, O. & Rotterová, P. (2018). Fuzzy Decision Matrices in Case of a Discrete Underlying Fuzzy Probability Measure. In J. Kacprzyk et al. (Eds.), *Advances in Fuzzy Logic and Technology 2017* (pp. 129–137). Heidelberg: Springer Internatioanal Publishing AG.

[7] Rotterová, P. & Pavlačka, O. (2016). Probabilities of Fuzzy Events and Their Use in Decision Matrices. *International Journal of Mathematics in Operational Research*, 9 (4), 423–435.

[8] Rotterová, P. & Pavlačka, O. (2017). New Approach to Fuzzy Decision Matrices. *Acta Polytechnica Hungarica*, 14 (5), 85–102.

[9] Rotterová, P. & Pavlačka, O. (2017). Decision matrices under risk with fuzzy states of the world and underlying discrete fuzzy probability measure. In P. Pražák (Ed.), *Proceedings of the 35th International Conference Mathematical Methods in Economics* (pp. 626–631). Hradec Králové: University of Hradec Králové.

[10] Sugeno, M. (1985). *Industrial applications of fuzzy control*. New York: Elsevier Science Pub. Co.

[11] Talašová, J. (2003). *Fuzzy methods of multiple criteria evaluation and decision making (in Czech)*. Olomouc: Publishing House of Palacký University.

[12] Talašová, J. & Pavlačka, O. (2006). Fuzzy Probability Spaces and Their Applications in Decision Making. *Austrian Journal of Statistics*, 35 (2&3), 347–356.

[13] Yoon, K.P. & Hwang, Ch. (1995). *Multiple Attribute Decision Making: An Introduction.* California: SAGE Publications.

[14] Viertl, R. (2011). *Statistical methods for fuzzy data*. Chichester: John Wiley & Sons, Ltd.

[15] Zadeh, L.A. (1968) Probability Measures of Fuzzy Events. *Journal of Mathematical Analysis and Applications*, 23 (2), 421–427.

[16] Zadeh, L.A. (1975) The Concept of a Linguistic Variable and its Application to Approximate Reasoning I. *Information Sciences*, 8, 199–245.

Matrix Rearrangement Problem with Partially Fixed Entries

Štefan Peško[1]

Abstract. In this paper, we study the new rearrangement problem of finding the intra-column permutation of a matrix such that the row-sums has minimal variability but some given elements are fixed. This problem occurs in assembly line scheduling problem, where n items are to be produced by m assembly lines. To each line we must assign d workers, specialized in different jobs. We have nd workers and each worker has a different job completion time. The objective is to maximize the probability of finishing of all n items in given deadline interval by assigning the workers to the assembly line.
For two columns matrix it is known that the standard rearrangement problem is polynomially solvable. For general matrices, this problem is NP-hard. We modify known stochastic Block Arrangement Algorithm. We study MILP and MIQP models for small number of fixed elements of matrix. This approach can be used as a heuristic method for medium size instances. Exact and heuristic solutions are discussed.

Keywords: matrix rearrangement, fixed elements, block arrangement algorithm, mixed integer programming.

JEL classification: 90C10, 90C15
AMS classification: C02, C61, C65, C68, J33

1 Introduction

We deal with solving a modified version of a NP-hard problem that occurs in assembly line crew scheduling problem, where n items are to be produced by m assembly lines. We study the new rearrangement problem of finding the intra-column permutation of a matrix such that the row-sums has minimal variability but some given elements are fixed.

First formulations under the name Matrix Permutation Problem (MPP) can be found in Černý [4] (in Slovak) and Tegze & Vlach [12] (in English). They are motivated by Peško's mathematical interpretation of a real problem, when it was needed to assign weekly work schedule to a given number of drivers as evenly as possible. The goal is to find a minimally varied set of row-sums of columns permuted matrix. The approaches based on graph theory solving graph version of the MPP are studied by Czimmermann [2]. In paper Peško and Černý [9] present different real situations where making managerial decisions can be influenced by the methodology of fair assignment.

For reference, 32 MPPs are defined and examined in work Dell'Olmoa [5]. They are characterized by particular measure of set uniformity to be optimized. 21 of the studied problems can be solved by linear time algorithms, 7 require more complex algorithms but can still be solved in polynomial time, and 3 are proved to be NP-hard. The possibility of use of algorithms based on aggregate two-column sub-problems for a fair scheduling can be found in papers Boud. Jakobsons & Vanduffel [1] or Puccetti & Ruschendorf [11]. The stochastic algorithm for uniform workload distribution problems is studied in paper Peško & Kaukič [10].

New interesting results can be found in paper Jakobsons and Wang [6] where arrangement increasing function with matrix input over intra-column permutations was studied. Authors show that many classical optimization problems, including stochastic crew scheduling and assembly of reliable systems, have objective functions with this structure. In the recent time the investigation of the problem continues, under the name of Matrix Arrangement Problem (MAP) or Block Rearranging Algorithm (BRA) in the papers Boud, Jakobsons & Vanduffel [1] and Puccetti, G. & Ruschendorg, L. [11].

Let us begin with motivation of our problem.

2 Assembly line crew scheduling problem

In the following subsections, we provide the modification of Assembly Line Crew Scheduling problem (ALCS) from paper Jakobsons, E. & Wang, R. [8] whose deterministic version was first introduced in Hsu [7]. We describe new stochastic version of this problem and we will use notation from [8].

[1]University of Žilina, Univerzitná 8215/1, 010 026 Žilina, Slovakia, stefan.pesko@fri.uniza.sk

The ALCS problem considers the production of n items on n parallel assembly lines (one item on each line), where each item requires d different operations. There are n workers specialized in each of the d operations, hence nd workers in total. The time taken by worker i of specialty j to complete the operation is a random variable $T_{ij} \sim F_j^{\theta_{ij}}$, where F_j^{θ} is a distribution function that models the completion time of operation j, parameterized by $\theta \in \Re$. All random variables $T_{ij}, i \in \mathbb{I} = \{1, \ldots, n\}, j \in \mathbb{J} = \{1, \ldots, d\}$ are assumed to be independent.

Thus, the input for an instance of this problem is the matrix of parameters $\boldsymbol{\Theta} \in \Re^{n \times d}$. We are allowed to assign the workers within each specialty to the n assembly lines in any order. If the i^{th} row of $\boldsymbol{\Theta}$ represents the i^{th} assembly line, then the assignment of workers corresponds to permuting the entries within each column (specialty). Denote by $C_i = \sum_{j=1}^{d} T_{ij}$ the *completion time* of the i^{th} item. Let c_L and c_U are given lower and upper bounds of completion times for all items.

Let $\mathcal{E} \subset \mathbb{I} \times \mathbb{J}$ be set of *fixed assignment* by some workers i of specialty j as additional input of the instance. This requirement is motivated by the practice that some workers are closely specialized in one type of operations. Our goal is maximize the probability of meeting completion times in interval $\langle c_L, c_U \rangle$ i.e.

$$\prod_{i \in \mathbb{I}} \mathcal{P}(c_L \leq C_i \leq c_U) \rightarrow \max. \tag{1}$$

Remark 1. Authors in paper [8] consider three different objectives for stochastic problem (maximize the probability of meeting a deadline, minimize the expected makespan and maximize the expected number of items finished within the deadline). They do not assume fixed assignments, only the makespan $C_{\max} = \max_{i \in \mathbb{I}} C_i < c_U$ of the project.

We will first formulate deterministic version of Matrix rearrangement problem with partially fixed entries.

3 Deterministic formulation

In this section we will use notation from section 2.

Let us have the real matrix $\boldsymbol{A} = (a_{ij}), i \in \mathbb{I} = \{1, 2, \ldots, n\}, j \in \mathbb{J} = \{1, 2, \ldots, d\}$ and the subset of fixed pair of indexes $\mathcal{E} \subset \mathbb{I} \times \mathbb{J}$. Let $\Pi(n, d)$ be set of all d-tuples of permutation of set $\{1, \ldots, n\}$. Then $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_d) \in \Pi(n, d)$ is vector of permutation elements of matrix $\boldsymbol{A^{\pi}} = (a_{\pi_j(i),j})$. Let $\boldsymbol{s^{\pi}} = (s_i^{\boldsymbol{\pi}}), i \in \mathbb{I}$ be a vector of row-sums, where $s_i^{\boldsymbol{\pi}} = \sum_{j \in \mathbb{J}} a_{\pi_j(i),j}$.

We consider two very simple object functions as an irregularity measures of vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$

1. $f_{dif}(\boldsymbol{x}) = \max_{i \in \mathbb{I}} x_i - \min_{i \in \mathbb{I}} x_i$,
2. $f_{var}(\boldsymbol{x}) = \sum_{i \in \mathbb{I}} (x_i - \overline{x})^2$, where $\overline{x} = \frac{1}{n} \sum_{i \in \mathbb{I}} x_i$.

Than aim of optimization problem FREA(f) is vector of permutations

$$\boldsymbol{\pi}^{*f} = argmin\{f(s^{\boldsymbol{\pi}}) : \boldsymbol{\pi} \in \Pi(n, d), \pi_j(i) = i \text{ for } (i, j) \in \mathcal{E}\},$$

where $f \in \{f_{dif}, f_{var}\}$.

Remark 2. It is possible to show that the feasible space of FREA(f_{var}) is a subset of the feasible space of FREA(f_{dif}).

3.1 Two-column case

For two-column case of the FREA(f), where $\boldsymbol{A} \in \Re^{n \times 2}$, the mean of row-sums equals to $\overline{a} = \frac{1}{n} \sum_{i \in \mathbb{I}} (a_{i1} + a_{i2})$. We can, without loss of generality, set $\pi_1(i) = i$ for $i \in \mathbb{I}$ and we get an objective function f_{var} in the form

$$\begin{aligned} f_{var}(s_1, s_2, \ldots, s_n) &= \sum_{i \in \mathbb{I}} (a_{i1} + a_{\pi_2(i)2} - \overline{a})^2 = \sum_{i \in \mathbb{I}} (a_{i1} + a_{\pi_2(i)2})^2 + \overline{a}^2 - 2\overline{a} \sum_{i \in \mathbb{I}} (a_{i1} + a_{\pi_2(i)2})) = \\ &= \sum_{i \in \mathbb{I}} (a_{i1}^2 + a_{\pi_2(i)2}^2 + 2a_{i1}a_{\pi_2(i)2}) + \overline{a}^2 - 2\overline{a} \sum_{i \in I} (a_{i1} + a_{i2}) = \\ &= 2 \sum_{i \in \mathbb{I}} a_{i1}a_{\pi_2(i)2} + \sum_{i \in I} (a_{i1}^2 + a_{i2}^2) + (1 - 2n)\overline{a}^2 = 2 \sum_{i \in \mathbb{I}} a_{i1}a_{\pi_2(i)2} + const. \end{aligned} \tag{2}$$

Thus, to minimize the quadratic function f_{var} it is sufficient to minimize the linear function $\sum_{i \in \mathbb{I}} a_{i1}a_{\pi_2(i)2}$ of the permutation π_2 of set $\mathbb{I}$.

Example 1. Let $\mathcal{E} = \{(3,1),(3,2),(6,1)\}$ be the set of fixed elements of matrix $\mathbb{A}$.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 2 & 3 \\ \mathbf{3} & \mathbf{2} \\ 5 & 4 \\ 6 & 6 \\ \mathbf{5} & 1 \end{pmatrix} \begin{matrix} 1 \\ 5 \\ 5 \\ 9 \\ 12 \\ 6 \end{matrix}, \boldsymbol{A^{\pi}} = \begin{pmatrix} 1 & 6 \\ 2 & 4 \\ \mathbf{3} & \mathbf{2} \\ 5 & 3 \\ 6 & 0 \\ \mathbf{5} & 1 \end{pmatrix} \begin{matrix} 7 \\ 6 \\ 5 \\ 8 \\ 6 \\ 6 \end{matrix}, \boldsymbol{\pi} = \begin{pmatrix} 1 & 5 \\ 2 & 4 \\ \mathbf{3} & \mathbf{3} \\ 4 & 2 \\ 5 & 1 \\ \mathbf{6} & 6 \end{pmatrix}$$

Note $\boldsymbol{s} = rowsums(\boldsymbol{A}) = (1,5,5,12,6)^T, \boldsymbol{s^{\pi}} = rowsums(\boldsymbol{A^{\pi}}) = (7,6,5,8,6,6)^T$. Then $f_{dif}(\boldsymbol{s})) = 11$, $f_{var}(\boldsymbol{s}) = 14.27$. Is $\boldsymbol{\pi}$ optimal solution with values of objective functions $f_{dif}(\boldsymbol{s^{\pi}})) = 2, f_{var}(\boldsymbol{s^{\pi}}) = 1.07$? ♠

From form (2) of objective function f_{var} we can see that the permutation π_2 we can define via binary variables x_{ik} for $(i,k) \in \mathbb{I} \times \mathbb{I}$:

$$x_{ik} = \begin{cases} 1 & \text{if } \pi_2(i) = k \\ 0 & \text{otherwise} \end{cases}$$

Now we can find optimal solution by solving the following linear assignment problem with reduction variables (LAP):

$$\sum_{(i,k)\in\mathbb{I}\times\mathbb{I}} a_{i1}a_{k2}x_{ik} \to \min \tag{3}$$

$$\sum_{i\in\mathbb{I}} x_{ik} = 1 \qquad \forall k \in I, \tag{4}$$

$$\sum_{k\in\mathbb{I}} x_{ik} = 1 \qquad \forall i \in I, \tag{5}$$

$$x_{ii} = 1 \qquad \forall (i,2) \in \mathcal{E}, \tag{6}$$

$$x_{ik} \geq 0 \qquad \forall i \in \mathbb{I}, \forall k \in \mathbb{I}. \tag{7}$$

The problem $(3) - (5), (7)$ is classical linear assignment problem. Constraint (6) ensures that in second column we chose fixed elements a_{i2} (first column is fixed implicitly) i.e. i^{th}-row and i^{th}-column of cost matrix can be deleted. It is known that the linear assignment problem of size $n_2 = n - |\mathcal{E}_2|, \mathcal{E}_2 = \{(i,2) \in \mathcal{E}\}$ is solvable in time $O(n_2^3)$ and so our two-column case is polynomially solvable.

But our special cost of assignment in objective function (3) enables us to use known procedure *quicksort* with the time complexity $O(n \log n)$. The following function $REA(\boldsymbol{A}, \mathcal{E})$ give optimal two-column matrix of permutations $\pi = (\pi_1, \pi_2)$ for rearrangement matrix $\mathbb{A}$ with fixed elements $\mathcal{E}$.

function $REA(\boldsymbol{A}, \mathcal{E})$:
  $E_0 = \{i \in I : (i,1),(i,2) \in \mathcal{E}\}$
  $E_1 = \{i \in I : (i,1) \in \mathcal{E}, (i,2) \notin \mathcal{E}\}$
  $E_2 = \{i \in I : (i,1) \notin \mathcal{E}, (i,2) \in \mathcal{E}\}$
  $P_1 = \text{argsort}(a_1, \downarrow), P_2 = \text{argsort}(a_2, \uparrow)$
  for $k \in \{1,2\}$:
    for $i \in E_0$:
      $j = P_k.index(i), P_k.remove(P_k(j)), P_k.insert(i,i)$
    for $i \in E_k$:
      $j = P_k.index(i), P_k.remove(P_k(j)), P_k.insert(i,i)$
  $\pi_1 = P_1, \pi_2 = P_2$
return (π_1, π_2)

Remark 3. If $\mathcal{E} = \emptyset$ then algorithm $REA(\boldsymbol{A}, \mathcal{E})$ minimize arrangement via oppositely ordered columns [8] with row-sum vector $\boldsymbol{s}^* = a_1^{\downarrow} + a_2^{\uparrow}$.

3.2 Multi-column case

For multi-column case of problem FREA(f), where $\boldsymbol{A} \in \Re^{n\times d}$ $d > 2$ the mean of row-sums is equal to $\overline{a} = \frac{1}{n}\sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{J}} a_{ij}$. For f_{dif} and f_{var} we can define the MILP an MILP model with binary variables x_{ijk} for $i,k \in \mathbb{I}, j \in \mathbb{J}$.

$$x_{ijk} = \begin{cases} 1 & \text{if } \pi_j(i) = k \\ 0 & \text{otherwise.} \end{cases}$$

The MILP model minimizes maximal difference between row-sums of matrix:

$$z_U - z_L \to \min, \tag{8}$$

$$\sum_{k\in\mathbb{I}} x_{ijk} = 1, \qquad \forall (i,j) \in \mathbb{I}\times\mathbb{J}, \tag{9}$$

$$\sum_{i\in\mathbb{I}} x_{ijk} = 1, \qquad \forall (k,j) \in \mathbb{I}\times\mathbb{J}, \tag{10}$$

$$x_{iji} = 1, \qquad \forall (i,j) \in \mathcal{E}, \tag{11}$$

$$z_L - \sum_{(k,j)\in\mathbb{I}\times\mathbb{J}} a_{kj}x_{ijk} \le 0, \qquad \forall i \in \mathbb{I}, \tag{12}$$

$$z_U - \sum_{(k,j)\in\mathbb{I}\times\mathbb{J}} a_{kj}x_{ijk} \ge 0, \qquad \forall i \in \mathbb{I}, \tag{13}$$

$$x_{ijk} \in \{0,1\}, \qquad \forall (i,j,k) \in \mathbb{I}\times\mathbb{J}\times\mathbb{I}, \tag{14}$$

$$z_L, z_U \ge 0. \tag{15}$$

The objective function f_{dif} is of the form (8). Constraints (9), (10) and (14) define feasible conditions for d-dimensional assignment problem. Constraints (11) guarrantee the fixation of elements of matrix. Constraints (12), (13) and (15) define lower and upper bounds of row-sums.

The MIQP model minimize variability of row-sums of matrix:

$$\sum_{i\in\mathbb{I}} (z_i - \bar{a})^2 \to \min, \tag{16}$$

$$z_i - \sum_{(k,j)\in\mathbb{I}\times\mathbb{J}} a_{kj}x_{ijk} = 0, \qquad \forall i \in \mathbb{I}, \tag{17}$$

$$x_{ijk} \in \{0,1\}, \text{ and } (9),(10),(11) \qquad \forall (i,j,k) \in \mathbb{I}\times\mathbb{J}\times\mathbb{I}, \tag{18}$$

$$z_i \ge 0, \qquad \forall i \in \mathbb{I}. \tag{19}$$

The objective function f_{var} has the form (16). Constraints (17) and (19) define i^{th} row-sum.

Models MILP and MIQP are NP-hard problems and also for medium problems we need a heuristic method. From different implementation of Block Rearrangement Algorithms (BREA) Boudt, Jakobsons & Vanduffel [1], Peško& Kaukič [10], Czimmermann, Peško & Černý [3] we modify following:

function $BREA(\boldsymbol{A}, \mathcal{E}, etime = 120)$:
  for $j \in J : \pi_j = (1, 2, \ldots, n)^T$
  $t = time()$
  while $time() - t < etime$:
    Choose randomly a non-empty subsets $J = J_1 \cup J_2, J_1 \cap J_2 = \emptyset$,
    $(\psi_1, \psi_2) = REA(\boldsymbol{B}, \mathcal{E}')$ with aggregated matrix $\boldsymbol{B} = (b_1, b_2)$,
      where $b_{i1} = \sum_{j\in J_1} a_{\pi_j(i),1}, b_{i2} = \sum_{j\in J_2} a_{\pi_j(i),2}$
      a $\mathcal{E}' = \{(i,k) : \exists (i,j) \in \mathcal{E}, j \in J_k, k \in \{1,2\}\}$.
    Sort elements of matrices $\boldsymbol{A}, \pi_k$ in columns J_k via ψ_k.
return $\boldsymbol{A}, \boldsymbol{\pi}$

Next we will show how use models MILP, MIQP and BREA method for solving one of stochastic versions of matrix rearrangement problems.

4 Stochastic formulation

Let us have the stochastic matrix $\boldsymbol{\mathcal{A}} = (A_{ij}), i \in \mathbb{I}, j \in \mathbb{J}$ where elements are independent random variables $A_{ij} \sim F_j^{\theta_{ij}}$ and F_j^{θ} is a distribution function. The fixed pair of indexes $\mathcal{E} \subset \mathbb{I}\times\mathbb{J}$ is given as in deterministic

formulation. Let $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_d) \in \Pi(n, d)$ be vector of permutation elements of matrix $\boldsymbol{\mathcal{A}^{\pi}} = (A_{\pi_j(i),j})$. Let $\boldsymbol{S^{\pi}} = (S_i^{\boldsymbol{\pi}}), i \in I$ be a random vector of row-sums, where $S_i^{\boldsymbol{\pi}} = \sum_{j \in J} A_{\pi_j(i),j}$.

The input for the instance of this problem are the matrix of parameters $\boldsymbol{\Theta} \in \Re^{n \times d}$ and real lower and upper bounds s_L and s_U. Our goal is to maximize the probability of row-sums of matrix in real interval $\langle s_L, s_U \rangle$ i.e.

$$\prod_{i \in \mathbb{I}} \mathcal{P}(s_L \leq S_i^{\pi} \leq s_U) \rightarrow \max, \tag{20}$$

$$\pi_j(i) = i \qquad \forall (i,j) \in \mathcal{E}, \tag{21}$$

$$\boldsymbol{\pi} \in \Pi(n, d) \tag{22}$$

We will call this problem *Stochastic Rearrangement Problem* (SREA) .

Remark 4. Real application of the SREA was formulated in motivation section 2 for ALCS problem where elements of matrices are $A_{ij} = T_{ij}$.

4.1 Two-column case

For two-column case of the SREA, where $\boldsymbol{\mathcal{A}} = (A_1, A_2), A_{ij} \sim F_{ij}(), i \in \mathbb{I}, j \in \{1, 2\}$, we search a matrix of permutations $\boldsymbol{\pi}^* = (\pi_1^*, \pi_2^*)$, such that

$$\boldsymbol{\pi}^* = argmax\{\log\ g(S^{\pi}) : \boldsymbol{\pi} \in \Pi(n, 2), \pi_j(i) = i \text{ for } (i,j) \in \mathcal{E}\},$$

where $g(\boldsymbol{\mathcal{S}^{\pi}}) = \prod_{i \in \mathbb{I}} \mathcal{P}(s_L \leq S_i^{\boldsymbol{\pi}} \leq s_U), S_i^{\boldsymbol{\pi}} = A_{i1} + A_{\pi_2(i),2}$ for $i \in \mathbb{I}$. Then we can write

$$\begin{aligned} \log\ g(\boldsymbol{\mathcal{S}^{\pi}}) &= \sum_{i \in \mathbb{I}} \log\ \mathcal{P}(s_L \leq A_{i1} + A_{\pi_2(i),2} \leq s_U) = \\ &= \sum_{i \in \mathbb{I}} \log\ (G_{\pi_2(i)}(s_U) - G_{\pi_2(i)}(s_L)), \end{aligned} \tag{23}$$

where $G_{\pi_2(i)}(s_.) = F_{i1} * F_{\pi_2(i),2}(s_.), i \in \mathbb{I}$ is a convolution of random variables A_{i1} a $A_{\pi_2(i),2}$.

Remark 5. Permutatiom π_1 here is the identity ($\pi_1(i) = i, i \in \mathbb{I}$) as in deterministic problem.

If we assume that matrix $\boldsymbol{\mathcal{A}}$ follow the Normal distribution i.e. $A_{ij} \sim \mathcal{N}(\theta_{ij}, \sigma_j^2), i \in \mathbb{I}, j \in \{1, 2\}$ then $F_{ij}(t) = \Phi\Big(\frac{t - \theta_{ij}}{\sigma_j}\Big)$. Denote $\theta_{i+} = \theta_{i1} + \theta_{i2}$ and $\sigma_+^2 = \sigma_1^2 + \sigma_2^2$. Then we can write $S_i \sim \mathcal{N}(\theta_{i+}, \sigma_+^2), i \in \mathbb{I}$. The postulate of homogeneity of variance σ_j^2 in column of matrix $\mathcal{A}$ allows us to consider only the matrix of mean values $\boldsymbol{\Theta} = (\theta_{ij}), i \in \mathbb{I}, j \in \{1, 2\}$.

Thus, it has been possible to transform this stochastic version of the problem into a deterministic i.e. then we can find solutions via rearrangement of matrix $\boldsymbol{\Theta}$. Solving MILP, MIQLP models resp BREA we get optimal resp. approximate solution $\boldsymbol{\pi}^*$ for which

$$\prod_{i \in \mathbb{I}} \Big[\Phi\Big(\frac{s_U - \theta_{i+}^*}{\sigma_+}\Big) - \Phi\Big(\frac{s_L - \theta_{i+}^*}{\sigma_+}\Big)\Big] \rightarrow maximum.$$

As we can see from equations (20) and (23) generalization to d-column case differs only by convolution d of random variables i.e. $\sum_{j=1}^{d} A_{\pi_j(i),j}$.

5 Computational experiments

Our experiments were conducted on PC Workstation (processor 8-core i7-5960X 3GHz, RAM 32GB) with OS Linux (Debian/stretch). We used Python-based tools and the Python interface to commercial mathematical programming solver Gurobi [6].

We experimented with 30 randomly generated instances of ALCS problem with parameters $n = 15, d = 4, A_{ij} \sim \mathcal{N}(\theta_{ij}, \sigma_j^2)$ where $\theta_{ij} \sim 5\sigma_j(1 + Beta(2, 5)), \sigma_j^2 = j$ for $i \in \mathbb{I}, j \in \mathbb{J}$ and $\sigma_+^2 = \sum_{j \in J} \sigma_j^2$, $\overline{\theta}_+ = \frac{1}{d} \sum_{i \in I} \sum_{j \in J} \theta_{ij}$ cited in [8] and $s_L = \overline{\theta}_+ - 3\sigma_+, s_U = \overline{\theta}_+ + 3\sigma_+$.

Results are summarized in table 1 for randomly fixed elements in number from 0 to 5 where we denote BREA–10, MILP–10, MIQP–10 the minimum and maximum probability of meeting completion times in first 10 executions and BREA–30, MILP–30, MIQP–30 from 30 executions. Execution time was limited to 120 seconds.

Table 1 Probability in meeting completion times in ALSC problem – range for firs 10 and 30 intanstances

	BREA	– 10	MILP	– 10	MIQP	– 10	BREA	– 30	MILP	– 30	MIQP	– 30
$\|\mathcal{E}\|$	min	max	min	max	min	max	min	max	min	max	min	max
0	0.	0.853	0.	0.957	0.	0.853	0.	0.853	0.	0.957	0.	0.853
1	0.087	0.837	0.875	0.955	0.187	0.837	0	0.899	0.	0.955	0.	0.899
2	0.	0.791	0.	0.95	0.	0.785	0.	0.909	0.	0.958	0.	0.909
3	0.	0.731	0.	0.955	0.	0.731	0	0.875	0	0.959	0	0.875
4	0.802	0.916	0.941	0.956	0.802	0.916	0	0.671	0	0.938	0	0.671
5	0.	0.558	0.	0.954	0.	0.741	0	0.875	0	0.957	0	0.875

6 Conclusion and future research

In this paper we studied new version of matrix rearrangement problem where partially fixed entries are given. Computation experiments has shown that the MILP, MIQP models and BREA for the ALCS problems gives applied heuristic solutions. Open question remains whether it is the exact method for the same spacial real instances. Note that the same approach would also be applicable if the workers on the same assembly line had dependent completion times then would be modeled by a multivariate normal distribution.

Acknowledgements

The research was supported by the research grants APVV-14-0658 Optimization of urban and regional public personal transport and VEGA 1/0582/16 Economic optimization of networks processes.

References

[1] Boudt, K., Jakobsons, E. & Vanduffel, S. (2017). Block Rearranging Elements within Matrix Columns to Minimize the Variability of the Row Sums, 40R, Forthcoming. Available at http://dx.doi.org/10.2139/ssrn.2634669

[2] Czimmermann, P.: A Note on Using Graphs in Regular Scheduling Problems, *Communications*, No.*4*, (2003), 47–48.

[3] Czimmermann, P., Peško, Š. & Černý, J. (2016). Uniform Workload Distribution Problems, *Communications*, no.*1A*, 56–59.

[4] Černý,J.: How to minimize irregularity? (Slovak), *Pokroky matematiky, fyziky a astronomie*, *30*(3), 1985, 145–150.

[5] DellOlmoa, P., Hansenb, P. & Pallottinoc, S. & Storchid G. (2005). On Uniform k-partition Problems, *Discrete Applied Mathematics*, Vol.*150*, Issues 13, 121–139.

[6] Gurobi Optimization Inc. *Gurobi optimizer reference manual*, 2015, http://www.gurobi.com.

[7] Hsu, W. L. (1984). Approximation algorithms for the assembly line crew scheduling problem, *Mathematics of Operations Research*, *9*(3), 376–383.

[8] Jakobsons, E. & Wang, R. (2016). Negative Dependence in Matrix Arrangement Problems (March 11, 2016). Available at SSRN: https://ssrn.com/abstract=2756934 or http://dx.doi.org/10.2139/ssrn.2756934

[9] Peško, Š. & Černý, J. (2006). Uniform Splitting in Managerial Decision Making, E+M, *Economics and Management IX*, Vol.4, Technical university of Liberec, 67–71.

[10] Peško, Š. & Kaukič, M. (2015). Stochastic Algorithm for Fair Workload Distribution Problem, In: *Proceedings of Informatics*, 13nd international scientific conference of informatics, Poprad, IEEE, pp. 206–210.

[11] Puccetti, G. & Ruschendorg, L. (2012) Computation and sharp bounds on expected value of a supermodular function of risk with given marginal. Computations in Statistics Simulation and Computation, 44(3), pp.705-718.

[12] Tegze, M. & Vlach, M. (1986). On the matrix permutation problem, *Zeitschrift fur Operations Research*, Vol. *30*, Issue 3, pp. A155–A159.

Portfolio selection model based on Drawdown performance measure

Juraj Pekár[1], Ivan Brezina[2], Ivan Brezina Jr.[3]

Abstract. The question of correct portfolio selection dominates always more in economics life. Correct portfolio selection allows fulfilling the investing goals. It does not matter whether the goal is the risk minimization, revenue maximization, or portfolio performance maximization. The model of portfolio selection is the tool, how to find the optimal portfolio composition depending on investors profile (risk attitude of investor). Many portfolio creation methods and portfolio management approaches were created. The main result of the presented model may not be only return maximization or risk minimization. It can be the maximization of performance, which the portfolio provides. The performance measure of investment presents results quantification, which we obtain by the chosen strategy of the investor. This quantification presents statistical summarization of reached revenue rate. It presents the estimation of investors risk and measure of manager skills of effective work with risk measure too. In this paper, we use performance measure of the portfolio based on Drawdown. The portfolio selection model is based on maximization of this Drawdown performance measure.

Keywords: portfolio performance rate, portfolio selection model, Drawdown

JEL Classification: C02, C61
AMS Classification: 90C11, 90B06

1 Introduction

The performance measure of investment ([3], [4]) is the quantification of achieved results by the chosen strategy of the investor. It is the statistical summarization of achieved returns measure, of possible risk estimation and of the investors ability, how effective the investor uses the information about the risk. We present here the theoretical optimization model of portfolio selection based on Drawdown performance measure. In this case, it is the one-criteria decision making because the performance measure is constructed on the base of the portfolio return and risk values.

Nowadays, is it normal practice that the investors split the investments on financial markets. To support decision making they use several calculations and analyses. The goal is to answer the question, how much can the investor lost in the selected period. Drawdown is the risk measure and the measure of investment strategy success ([1], [2], [5]). On investment market is Drawdown defined as the biggest decline of portfolio return.

The maximum value of Drawdown measures highest percentage decline in portfolio value in selected period. This value can be used as the measure of the portfolio risk. It is the indicator of the portfolio management quality. The value of Drawdown depends on the maximum and minimum value of asset in the test period. It gives investor the information about investment strategy success, about the investment strategy risk and about the finance, which investor needs to realize the investment. The value of Drawdown is the indicator of portfolio risk and investment strategy success. The low value of Drawdown indicates good portfolio management and small risk of selected portfolio.

Calmar ratio, Sharpe ratio, Omega function, Sortino ratio are the often-used performance rate of the portfolio ([6]). The portfolio selection model based on Drawdown performance rate is a new approach, which is based on Sharpe ratio. The difference between this models is, that the portfolio selection model based on Drawdown uses Drawdown value instead of classic statistical dispersion.

[1] 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.
[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 Economics, Faculty of Economics, Pan-European university, Tematínska 10, 851 05 Bratislava, e-mail: brezina.ivan@yahoo.com

2 Performance rate based on Drawdown

In this section, we analyse performance rate measure based on Drawdown. As first we describe Calmar ratio as selected performance rate. The Calmar Ratio is an important statistic used to measure return vs. Drawdown risk ([7]). The Calmar ratio is the comparison of the estimated rate of return and of maximum risk level on the finance market. The lower is the value of Calmar ratio, the worse is the investment in selected period. The higher is the value of Calmar ratio, the better is the investment in the selected period.

The performance measures based on Sharpe ratio are in many publications ([6]). The Sharpe ratio is the average return earned in excess of the risk-free rate per unit of volatility or total risk. Drawdown risk measure instead of statistical dispersion is used in presented model.

Let X be a continuous random variable in time $t \in \langle 0,T \rangle$ with values X_t. Then the Drawdown value in T is:

$$DD(X) = \max_{t \in \langle 0,T \rangle} (X_t - X_T). \tag{1}$$

DD(X) presents maximum difference between maximum return in the time from first to last observation and present value of return.

Because the Calmar ratio is based on maximum Drawdown, in next section will be presented the calculation of maximum Drawdown. The value of maximum Drawdown is calculated by (1). The maximum Drawdown represents the biggest decline of return in whole period. For continuous random variable X the values are defined as follows:

$$MDD(X) = \max_{t \in \langle 0,T \rangle} \left\{ \max_{u \in \langle 0,t \rangle} (X_u - X_t) \right\}. \tag{2}$$

Sharpe ratio is the base of Calmar ratio. Young [7] created here presented performance measure. In this performance measure the maximum Drawdown value is used and not the standard deviation, which is used in Sharpe ratio. Formulation of this performance measure is follows:

$$C_P(X) = \frac{E(X) - r_f}{MDD(X)} = \frac{E(X) - r_f}{\max_{t \in \langle 0,T \rangle} \left\{ \max_{u \in \langle 0,t \rangle} (X_u - X_t) \right\}} \tag{3}$$

If we consider the existence of continuous random variable, then the formulation of presented performance measure is as follows. Let **r** be a vector of discrete random variable X. X is the measure of the portfolio return in T periods. The expected return of this investment at the same confidence level of all values of vector **r** is:

$$E(\mathbf{r}) = \frac{1}{T} \sum_{t=1}^{T} r_t \tag{4}$$

The value of maximum Drawdown for discrete random variable based on (2) is:

$$MDD(\mathbf{r}) = \max_{t=1,2,\ldots T} \left\{ \max_{u=1,2,\ldots t} (r_u - r_t) \right\} \tag{5}$$

Than the portfolio performance can be defined as the division of the asset return reduced by a risk-free value r_f by maximum Drawdown value of discrete random variable.

$$C_P(\mathbf{r}) = \frac{E(\mathbf{r}) - r_f}{MDD(\mathbf{r})} = \frac{1/T \sum_{t=1}^{T} r_t - r_f}{\max_{t=1,2,\ldots T} \left\{ \max_{u=1,2,\ldots t} (r_u - r_t) \right\}} \tag{6}$$

3 Portfolio selection model based on Drawdown performance rate

The performance rate $G(X)$ is the division of expected return by relative risk measure. The goal of investor is to maximize this performance rate. Let A be a set of X random variables. These random variables are the returns of portfolios, which can be constructed. The $\rho(X)$ is the function of risk. The $V(X)$ is the function of expected return. The problem of portfolio performance maximization is the problem of portfolio selection model:

$$\begin{aligned} &\max G(X) = \frac{V(X)}{\rho(X)} \\ &X \in A \end{aligned} \tag{7}$$

In this section, we formulate the portfolio selection model. In this model we use discrete random variables. We consider to construct the portfolio of n assets, which vectors of return are $\mathbf{r}_1$, $\mathbf{r}_2$,...$\mathbf{r}_n$. $\mathbf{r}_1$, $\mathbf{r}_2$,...$\mathbf{r}_n$ and are discrete random variables. We assume, that the investor invests in various assets. The weights of these assets on the portfolio are $\mathbf{w} = (w_1, w_2, ...w_n)^{\mathrm{T}}$. Let V denotes the function of the expected return. Let ρ denotes the function of the portfolio risk. Then the mathematical programming model based on G maximization can be formulated as follows:

$$\begin{aligned} &\max G(w_1, w_2, ...w_n) = \frac{V(w_1, w_2, ...w_n)}{\rho(w_1, w_2, ...w_n)} \\ &\sum_{i=1}^{n} w_i = 1 \\ &w_1, w_2, ...w_n \geq 0 \end{aligned} \tag{8}$$

The optimal weights of assets in the portfolio are the result of this model, which maximize the performance function. In next section we construct the portfolio selection model, which consider Calmar ratio. The goal of the portfolio selection model is to maximize this Calmar ratio.

Let r_{jt} be t-th value of discrete random variables X_j. Let w_j be a proportion of j-th asset in investment j=1, 2, ..., n. Than the expected return of this investment in selected period is:

$$E(w_1, w_2, ..., w_n) = E\left[\sum_{j=1}^{n} \mathbf{r}_j w_j\right] = \sum_{j=1}^{n} w_j E\left[\mathbf{r}_j\right] \tag{9}$$

The function of maximum Drawdown is defined by the equation (5). The modification of this equation for the portfolio with weights $\mathbf{w} = (w_1, w_2, ...w_n)^{\mathrm{T}}$ is as follows:

$$MDD(\mathbf{w}) = \max_{t\in(0,T)} \left\{ \max_{u\in(0,t)} \left(\sum_{j=1}^{n} {r_{ju}}^{\mathrm{T}} w_j - \sum_{j=1}^{n} {r_{jt}}^{\mathrm{T}} w_j \right) \right\} \tag{10}$$

In previous section Calmar ratio (6) was defined. The problem of portfolio selection can be formulated as the mathematic programming model based on this Calmar ratio. The result of this model is the optimal portfolio with maximum Calmar Ratio. The model is following:

$$\begin{aligned} &\max C_P(w_1, w_2, ...w_n) = \frac{\sum_{j=1}^{n} w_j E\left[\mathbf{r}_j\right] - r_f}{\max_{t=1,2,...T} \left\{ \max_{u=1,2,...t} \left(\sum_{j=1}^{n} {r_{ju}}^{\mathrm{T}} w_j - \sum_{j=1}^{n} {r_{jt}}^{\mathrm{T}} w_j \right) \right\}} \\ &\sum_{i=1}^{n} w_i = 1 \\ &w_1, w_2, ...w_n \geq 0 \end{aligned} \tag{11}$$

This model has to be transformed in order to get the linear function of risk, which represents maximum Drawdown. We need to transform $\max_{t=1,2,...T}\left\{\max_{u=1,2,...t}\left(\sum_{j=1}^{n} r_{ju}{}^{\mathrm{T}} w_j - \sum_{j=1}^{n} r_{jt}{}^{\mathrm{T}} w_j\right)\right\}$. For this transformation ([1]) we use the variables u_t, where $t = 1, 2,...T$ and $u_0=0$. The value of these variables is the maximum value of return to state t. Conditions (14) – (16) provide the required values of variables u_t. The next step is to define variable α, which has the value of the maximum Drawdown. Variable α has the value of maximum Drawdown because of constrain (13) and because the variable α is the denominator of the objective function, that is maximized. The mathematical programming model is than as follows:

$$\max C_P\left(w_1, w_2, ...w_n, u_1, u_2, ...u_T, \alpha\right) = \frac{\sum_{j=1}^{n} w_j E\left[\mathbf{r}_j\right] - r_f}{\alpha} \tag{12}$$

$$u_t - \sum_{j=1}^{n} r_{jt}{}^{\mathrm{T}} w_j - \alpha \le 0, t = 1, 2, ...T \tag{13}$$

$$u_t \ge \sum_{j=1}^{n} r_{jt}{}^{\mathrm{T}} w_j, t = 1, 2, ...T \tag{14}$$

$$u_t \ge u_{t-1}, t = 1, 2, ...T \tag{15}$$

$$u_0 = 0 \tag{16}$$

$$\sum_{i=1}^{n} w_i = 1 \tag{17}$$

$$w_1, w_2, ...w_n \ge 0 \tag{18}$$

The denominator can't be equal to zero. It makes the problem unsolvable. In this model, it can happen only when the portfolio return will never decrease. It is impossible, that returns will be always decreasing. It means that problem, which is solved in previous model (12) – (18) is always solvable.

We formulate the problem in software using nonlinear model (12) – (18). Model is implemented in the software GAMS (solver CONOPT) on PC with Intel ® Core ™ i7-3770 CPU with a frequency of 3.40 GHz and 8 GB of RAM under MS Windows 8 (Figure 1).

```
Sets t
sub(t)
i ;
Scalar RF ;
Table r(t,i) ;
Variables CR,x,alfa,u(t);
Positive variables w(i);
Equations
ucel
ohr
ohrx(i)
ohr0
ohr1(sub)
ohr2(t)
ohr3(t)
ohr4;
ohr.. sum(i,w(i))=e=1;
ohrx(i).. w(i)=l=1;
ohr0..x=e=sum((i,sub),w(i)*r(sub,i))/T-RF;
ohr1(sub)..alfa=g=u(sub)-sum(i,r(sub,i)*w(i));
ohr2(sub)..u(sub)=g=sum(i,r(sub,i)*w(i));
ohr3(sub)..u(sub)=g=u(sub-1);
ohr4..u('0')=e=0;
ucel.. CR=e=x/alfa;
Model CRF /all/
Solve CRF using dnlp maximizing CR;
```

Figure 1 GAMS code

The result is the solution of this problem, which consist of asset proportions of invested capital value in the optimal portfolio.

4 Conclusion

Many authors deal with above presented problematic. They construct several types of performance measure, which meet the requirements of practice for simplicity of calculation and for economic implementation too. On the other hand, the functions with presented measures should meet the properties, which help to make the presented problems easy to solve. Nowadays it is the main goal to find suitable performance measure, which can be generally accepted.

Presented approach in this paper can be used in the portfolio selection process. This portfolio selection can be realized by presented portfolio selection model. Presented portfolio selection model based on Drawdown performance measure is a new original approach, which originates from Sharpe ratio, but uses Drawdown risk measure instead of classic dispersion. The whole process of portfolio selection is very complicated and this paper presents one field affecting the final investor decision.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic – VEGA, grant no. No. 1/0351/17 Application of selected models of game theory to solve some economic problems of Slovakia

References

[1] Cheklov, A., Uryasev, S. and Zabarankin, M. (2003). Portfolio Optimization with Drawdown Constraint. In: Scherer B (Ed.): *Asset and Liability Management Tools* (pp. 263–278). London: Risk Books.
[2] Grossman, S. J. and Zhou, Z. (1993). Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3, 241–276 .
[3] Filbeck, G. and Tompkins, D. L. (2004). Management Tenure and Risk-Adjusted Performance of Mutual Funds. *Journal of Investing*, 13, 72–80.
[4] Keating, C. – Shadwick, W. F. (2002). A Universal Performance Measure. *Journal of Performance Measurement*, .
[5] Le Sourd, V. (2007). *Performance Measurment Traditional Investment*. Literature Survey. Working paper, E D H E C Risk and Asset Management Research Center, Lille-Nice.
[6] Pekár, J., Čičková, Z. and Brezina, I. (2016). Portfolio performance measurement using differential evolution. *Central European Journal of Operations Research*, 24, 421-433.
[7] Young, T.W. (1991). Calmar Ratio: A Smoother Tool. *Futures*, 20, 40.

Three-dimensional rectangular packing problem

Jan Pelikán[1], Jan Fábry[2]

Abstract. At present, the vast majority of transport flows is carried out by means of containers, trucks, wagons, pallets etc. In this sense, there is a significant increase in requirements for effective use of the cargo space, i.e. the packing that leads to the minimization of the unused space or of a number of used containers. The contribution is concerned with the three-dimensional packing problem in which rectangular boxes must be effectively placed into containers. In this paper we review the research focusing on the mathematical models providing the exact solution and heuristic algorithms giving approximate solution. Real data is used to verify presented tools. Based on numerical experiments, we show the limitations of mathematical models due to the large number of binary variables and corresponding constraints. Therefore, heuristic algorithms are used in real applications.

Keywords: three-dimensional packing, container loading, logistics, heuristics.

JEL classification: C44
AMS classification: 90C15

1 Introduction

In logistics, packing problems are defined as the task of placement of small items in two- or three-dimensional large objects. In case of three dimensions (3D), containers or trucks are used for transport of items that can be placed directly or with the use of pallets. In general, small items may have rectangular or irregular shape. The paper is aimed at the 3D packing problem of transport rectangular items, i.e. boxes. Because heterogeneous boxes are mostly transported in real instances, the optimization problem is strongly difficult to be solved. Therefore, heuristic algorithms need to be developed for finding good solutions in available time. The problem becomes even more complicated if heterogeneous set of containers or heterogeneous fleet of trucks are used. From this point of view, the following classification of problems is given [4]: identical item packing problem, single large object placement problem, single knapsack problem, multiple identical large object placement problem, multiple heterogeneous large object placement problem, multiple identical knapsack problem and multiple heterogeneous knapsack problem.

In all 3D problems the objective is obviously the maximization of the container volume utilization, or alternatively the minimization of a number of containers used for transport of the items. In the paper, we assume that all boxes can only be placed with their edges parallel to the floor and the walls of containers. Bortfeldt and Wäscher make a review of constraints that may be respected in container loading [1]. In many cases the total weight of loaded boxes may be more restrictive limitation than volume constraints. In addition, a convenient distribution of weight of the cargo across the container floor (load balance) is often considered. In real instances, there are given loading priorities or requirements corresponding to the order in which items will be later unloaded. When the rectangular items are transported, three vertical orientations are possible. The orientation of many boxes is restricted because of special types of goods placed inside. Of course, the horizontal orientation may be set in advance as well. If boxes are stacked up in a container, some of them cannot be placed in lower levels because of their possible damaging. Those limitations are often connected with stability requirements. Allocation constraints can be considered in multiple container problems. They are associated with the necessity to deliver boxes to the same destination. Therefore, such boxes must be located in the same container.

2 Mathematical Model

For rectangular packing problem, the following mathematical model, based on the model in [2], can be formulated. Let N be a total number of boxes to be packed. For each box, three parameters are given: length p_i, width q_i and height r_i $(i = 1, 2, \ldots, N)$. Let m be a total number of containers available. Their parameters are: length L_j, width W_j and height H_j $(j = 1, 2, \ldots, m)$. We define M as a large number.

Let us introduce binary variable s_{ij} which is equal to 1 if a box i is placed to container j, 0 otherwise. Binary variable n_j is equal to 1 if container j is used, 0 otherwise. For the location of box i in container there are defined

[1]University of Economics Prague, Department of Econometrics, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, pelikan@vse.cz
[2]University of Economics Prague, Department of Econometrics, W. Churchill sq. 4, 130 67 Prague 3, Czech Republic, fabry@vse.cz

three non-negative variables: x_i, y_i and z_i. They refer to the front-left bottom corner of the box. Binary variables lx_i, wx_i and hx_i indicate whether the length, width or height of box i is parallel to X-axis. Similarly, variables ly_i, wy_i and hy_i are defined for Y-axis and variables lz_i, wz_i and hz_i are defined for Z-axis. Finally, binary variables pl_{ik}, pr_{ik}, pbh_{ik}, pf_{ik}, pbl_{ik} and pa_{ik} are defined for the indication of the relative positions of boxes i and k ($i < k$). Variable pl_{ik} is equal to 1 if box i is placed on the left side of box k, 0 otherwise. Similarly, positions are right of, behind, in front of, below and above. We assume that each container is oriented with its length along X-axis and its width along Y-axis. The origin of the coordinate system represents front-left bottom corner of each container.

The mathematical model of the problem is then formulated as follows:

$$\text{Minimize} \qquad \sum_{j=1}^{m} L_j W_j H_j n_j - \sum_{i=1}^{N} p_i q_i r_i \tag{1}$$

$$x_i + p_i lx_i + q_i wx_i + r_i hx_i \leq x_k + (1 - pl_{ik})M \quad \text{for all } i,k,\ i < k, \tag{2}$$

$$x_k + p_k lx_k + q_k wx_k + r_k hx_k \leq x_i + (1 - pr_{ik})M \quad \text{for all } i,k,\ i < k, \tag{3}$$

$$y_i + q_i wy_i + p_i ly_i + r_i hy_i \leq y_k + (1 - pbh_{ik})M \quad \text{for all } i,k,\ i < k, \tag{4}$$

$$y_k + q_k wy_k + p_k ly_k + r_k hy_k \leq y_i + (1 - pf_{ik})M \quad \text{for all } i,k,\ i < k, \tag{5}$$

$$z_i + r_i hz_i + q_i wz_i + p_i lz_i \leq z_k + (1 - pbl_{ik})M \quad \text{for all } i,k,\ i < k, \tag{6}$$

$$z_k + r_k hz_k + q_k wz_k + p_k lz_k \leq z_i + (1 - pa_{ik})M \quad \text{for all } i,k,\ i < k, \tag{7}$$

$$pl_{ik} + pr_{ik} + pbh_{ik} + pf_{ik} + pbl_{ik} + pa_{ik} \geq s_{ij} + s_{kj} - 1 \quad \text{for all } i,k,j,\ i < k, \tag{8}$$

$$x_i + p_i lx_i + q_i wx_i + r_i hx_i \leq L_j + (1 - s_{ij})M \quad \text{for all } i,j, \tag{9}$$

$$y_i + q_i wy_i + p_i ly_i + r_i hy_i \leq W_j + (1 - s_{ij})M \quad \text{for all } i,j, \tag{10}$$

$$z_i + r_i hz_i + q_i wz_i + p_i lz_i \leq H_j + (1 - s_{ij})M \quad \text{for all } i,j, \tag{11}$$

$$\sum_{j=1}^{m} s_{ij} = 1 \quad \text{for all } i, \tag{12}$$

$$\sum_{i=1}^{N} s_{ij} \leq M n_j \quad \text{for all } j, \tag{13}$$

$$lx_i, ly_i, lz_i, wx_i, wy_i, wz_i, hx_i, hy_i, hz_i \text{ are binary for all } i, \tag{14}$$

$$pl_{ik}, pr_{ik}, pbh_{ik}, pf_{ik}, pbl_{ik}, pa_{ik} \text{ are binary for all } i,k, \tag{15}$$

$$s_{ij} \text{ is binary for all } i,j, \tag{16}$$

$$n_j \text{ is binary for all } j, \tag{17}$$

$$x_i, y_i, z_i \geq 0 \text{ for all } i. \tag{18}$$

The objective (1) is to minimize the total unused space of selected containers. Constraints (2)–(7) ensure that boxes do not overlap each other. Inequalities (8) are related to two different boxes i and k. If both boxes are located in different containers, $s_{ij} + s_{kj} - 1$ is equal to 0 or -1, and the constraint is always valid because there is the sum of binary variables on the left-hand side. On the contrary, if boxes i and k are located in the same container, $s_{ij} + s_{kj} - 1$ is equal to 1, and the constraint determines that boxes are placed behind each other, side by side or above each other, i.e. at least one variable on the left-hand side is equal to 1. Such variable affects the corresponding right-hand sides of constraints (2)–(7) and also coordinates of boxes location x_i, x_k, y_i, y_k, z_i and z_k. Respecting constraints (9)–(11), placement of all boxes satisfies the size of containers. The set of equations (12) ensures that each box will be placed in exactly one container. If at least one box is placed in a container, this container is used, and on the contrary, if a container is not used, no box can be loaded into it. These conditions are assured by constraints (13).

Gencer modified the model introducing additional index to the variables for the purpose of easier identifying the container in which the box is placed [3]. The main difference between this modification and model (1)–(18) consists in the substitution of variables x_i, y_i and z_i, determining the location of box i in a container, with variables x_{ij}, y_{ij} and z_{ij}, where index j identifies the container in which box i is located. This provides the immediate information about the location of boxes in containers. However, if box i is not located in container j, then variables x_{ij}, y_{ij} and z_{ij} are equal to 0, which is misleading, because these values correspond to coordinates (0, 0, 0) of the front-left bottom corner of box i in container j, while zero value of binary variable s_{ij} determines that box i is not in container j. Simultaneously, the Gencer's modification leads to significant increase in a number of variables and constraints. Table 1 shows the numbers of variables and constraints in both models depending on the size of the instance. In the following chapter we verify both models on real cases and provide their comparison.

	Chen et al.	Gencer
ALL variables	$N(3N+m+4)+m$	$N(mN+4m+1)$
BIN variables	$N(3N+m+1)+m$	$N(mN+m+1)$
Constraints	$N(10N-6)+m$	$N(7m(N-1)+4)+6N+2m$

Table 1 Model size

3 Application of the Model

We applied the mathematical model defined in the previous chapter on two cases. In the first illustrative example, 9 boxes are to be loaded in 2 available containers. In the second real instance, there is 52 boxes that must be loaded in 4 containers. In Table 2 there are given dimensions (in cm) of each box to be loaded. The length, width and height of containers are 350, 200 and 200 cm. As the results, table shows the coordinates x, y and z of front-left bottom corner of each box and its orientation (X means that the length of the box is parallel to the length of the container, Y means that its width is parallel to the length of the container). Because boxes are stacked up in two levels, we identify in which level of a container the box is placed. It is obvious that only one container is used for the transportation of boxes. Figures 1 and 2 show the floor schedule of placement of boxes in the container.

Box	Length	Width	Height	x	y	z	Level	Orientation
1	132	96	162	254	0	0	lower	Y
2	120	80	40	0	0	120	upper	X
3	120	80	70	134	0	0	upper	X
4	132	96	162	122	0	104	lower	X
5	122	102	110	132	0	0	lower	X
6	120	80	82	0	0	0	lower	X
7	122	102	110	0	0	0	upper	X
8	53	49	52	0	0	80	lower	X
9	120	60	113	0	0	140	lower	X

Table 2 Placement of 9 boxes

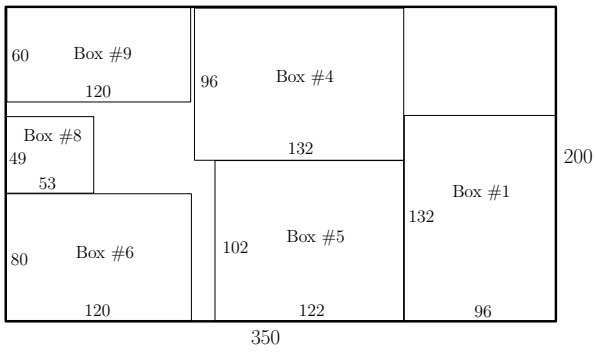


Figure 1 Lower level schedule

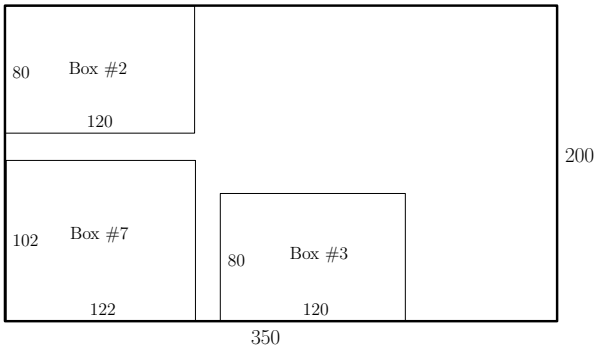


Figure 2 Upper level schedule

The real instance deals with the load of 52 boxes. For this purpose, we have 4 containers at disposal. Table 3 compares the results of Chen's et al. model and its Gencer's modification in terms of the computational complexity. Calculations were executed in CPLEX 12.5 on the computer with Intel Core i7-3610QMU CPU 2.30 GHz. As results show, model (1)–(18) is more suitable for finding the optimum solution to real cases. The information about the assignment of box i to container j is contained in variable s_{ij}. It is equal 1 if box i is located in container j, 0 otherwise. Exact location of the box inside the container is given by values of variables x_i, y_i and z_i.

4 Conclusions

The paper deals with the placement of 3D rectangular items in containers. The objective is to minimize the number of containers used. For this purpose, two types of mathematical models were applied on data from case study in two variants. The calculation experiment was used to verify models on the first variant with a small number of boxes. We present the optimal placement of boxes in two levels of the container (see Figures 1 and 2). The

	Chen et al.		Gencer	
Boxes	9	52	9	52
ALL variables	299	8532	542	32868
BIN variables	272	8376	488	32244
Constraints	391	14152	679	38852
Time (sec)	< 1	64	< 1	1202

Table 3 Computational results

second experiment contained a real list of 52 boxes that had to be assigned to the minimum number of containers. In this experiment, the results of both models were determined and compared as to the objective value and as to the calculation time. In conclusion, the application of model (1)–(18) on the case study data gives an efficient solution in significantly better calculation time than the run of the second model. Increasing a number of boxes to 103, the first model gave the optimum solution in approximately 2 hours, while the Gencer's modification did not provide a solution because of the frozen run. The studied 3D problem is topical due to the large amount of goods transported in trucks and other means of transport. The described model brings real savings not only in costs but also in reducing negative environmental impacts.

Acknowledgements

This paper was supported by the grant No. 16-00408S of the Czech Grant Agency.

References

[1] Bortfeldt, A. and Wäscher, G. (2013). Constraints in container loading – A state-of-the-art review. *European Journal of Operational Research*, 229, 1–20.

[2] Chen, C. S., Lee, S. M. and Shen, Q. S. (1995). An analytical model for the container loading problem. *European Journal of Operational Research*, 80, 68–76.

[3] Gencer, C. (2001). Modification of an analytical model for container loading problems. *AMSE, Modelling, Measurement and Control D.* 22, 51–58.

[4] Zhao, X., Bennell, J. A., Bektaş, T. and Dowsland, K. (2016). A comparative review of 3D container loading algorithms. *International Transactions in Operational Research*, 23, 287–320.

Japanese candle for financial portfolio

Krzysztof Piasecki[1], Anna Łyczkowska-Hanćkowiak[2]

Abstract. In this paper imprecise present values of securities are modeled by Japanese candles which are described by ordered fuzzy numbers. Thanks to this approach, we show that the imprecise present value of portfolio is presented as linear combination of Japanese candles of portfolio components. The possibility of applying this result in the portfolio analysis was indicated.

Keywords: Japanese candle, ordered fuzzy number, portfolio, present value

JEL Classification: C44, C02, G10
AMS Classification: 03E72, 91G10

1 Introduction

Japanese candles [5] are very useful tool for analysing a single security. There are also widespread analyses of Japanese candlesticks with definitions as short time series of Japanese candles for fixed single value paper. In [3], Japanese candles are described by ordered fuzzy numbers [2, 9]. In this paper we propose some way of Japanese candles application for portfolio analysis.

For any security, its imprecise present value may be determined as its Japanese candle. We will focus our attention on the problem of the use of Japanese candles to determine the present value of the securities' portfolio. The main goal of this paper is to find a relationship between portfolio present value and Japanese candles of portfolio components. This is a necessary starting point for more complex portfolio analyses. The directions for further research will be indicated.

2 Ordered fuzzy numbers

The concept of ordered fuzzy numbers (OFN) was intuitively introduced by Kosiński et al. [2] as an extension of the concept of fuzzy number (FN) [1]. A significant drawback of Kosiński's theory is that there exist such OFNs which, in fact, are not FN. What is more, the intuitive Kosiński's approach to the notion of OFN is very useful. The OFNs' usefulness follows from the fact that an OFN is a FN supplemented by orientation $[\![a \rightarrowtail d]\!] = (a, d)$ which is understood as a linear order of real numbers. For this reason, the Kosiński's theory was revised in [9] where OFNs are generally defined in following way:

Definition 1: *Let for any nondecreasing sequence* $(a, b, c, d) \subset \mathbb{R}$ *the starting-function* $S_{\mathcal{L}}: [a, b] \longrightarrow [0,1]$ *and the ending-function* $E_{\mathcal{L}}: [c, d] \longrightarrow [0,1]$ *are upper semi-continuous monotonic functions satisfying the condition*

$$S_{\mathcal{L}}(b) = E_{\mathcal{L}}(c) = 1. \tag{1}$$

Then ordered fuzzy number $\overleftrightarrow{\mathcal{L}}(a, b, c, d, S_{\mathcal{L}}, E_{\mathcal{L}})$ *is defined as the pair of fuzzy number determined by its membership function* $\mu_{\overleftrightarrow{OFN}}(\cdot\,|a, b, c, d) \in [0; 1]^{\mathbb{R}}$ *given by the identity*

$$\mu_{\overleftrightarrow{OFN}}(x|a, b, c, d, S_{\mathcal{L}}, E_{\mathcal{L}}) = \begin{cases} 0, & x \notin [a, d] = [d, a], \\ S_{\mathcal{L}}(x), & x \in [a, b] = [b, a], \\ 1, & x \in [b, c] = [c, b], \\ E_{\mathcal{L}}(x), & x \in [c, d] = [d, c] \end{cases} \tag{2}[3]$$

and orientation $[\![a \rightarrowtail d]\!] = (a, d)$.□

The space of all OFN is denoted by the symbol $\mathbb{K}$. For the case $a > d$ the orientation $[\![a \rightarrowtail d]\!]$ is negative. The interpretation of negatively oriented number is that is a number that may decrease. The space of all negatively oriented OFN is denoted by the symbol $\mathbb{K}^{-}$. For the case $a < d$ the orientation $[\![a \rightarrowtail d]\!]$ is positive. The interpretation of positively oriented number is that is a number that may increase. The space of all positively oriented OFN

[1] Poznań University of Economics, Department of Investment and Real Estate, al. Niepodległości 10, 61-875 Poznań, Poland, E-mail: krzysztof.piasecki@ue.poznan.pl
[2] WSB University in Poznań, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland, e-mail:anna.lyczkowska-hanckowiak@wsb.poznan.pl
[3] Let us note that this identity describes additionally extended notation of numerical intervals, which is used in this work.

is denoted by the symbol $\mathbb{K}^+$. For the case $a = d$, OFN $\overrightarrow{\mathcal{L}}(a, a, a, a, S_{\mathcal{L}}, E_{\mathcal{L}})$ represents crisp number $a \in \mathbb{R}$, which is not oriented. The space $\mathbb{K}$ may be described as the union

$$\mathbb{K} = \mathbb{K}^+ \cup \mathbb{K}^- \cup \mathbb{R}, \tag{3}$$

In this paper, we will limit our considerations to the following kind of OFN.

Definition 1.[9] *For any monotonic sequence* $(a, b, c, d) \subset \mathbb{R}$ *the trapezoidal ordered fuzzy number (TrOFN)* $\overrightarrow{Tr}(a, b, c, d)$ *is defined as the pair of FN determined by its membership function* $\mu_{\overrightarrow{Tr}}(\cdot\,|a, b, c, d) \in [0; 1]^{\mathbb{R}}$ *given by the identity*

$$\mu_{\overrightarrow{Tr}}(x|a, b, c, d) = \begin{cases} 0, & x \notin [a, d] = [d, a], \\ \frac{x-a}{b-a}, & x \in [a, b[=]b, a], \\ 1, & x \in [b, c] = [c, b], \\ \frac{x-d}{c-d}, & x \in]c, d] = [d, c[\end{cases} \tag{4}$$

and orientation $[\![a \rightarrowtail d]\!]$. □

The condition $a < d$ fulfilment determines the positive orientation $[\![a \rightarrowtail d]\!]$ of TrOFN $\overrightarrow{Tr}(a, b, c, d)$. The condition $a > d$ fulfilment determines the negative orientation $[\![a \rightarrowtail d]\!]$ of TrOFN $\overrightarrow{Tr}(a, b, c, d)$. For the case $a = d$, TrOFN $\overrightarrow{Tr}(a, a, a, a)$ represents crisp number $a \in \mathbb{R}$, which is not oriented.

In agreement with arithmetic introduced in [9], for the case of any TrOFNs $\overrightarrow{Tr}(a, b, c, d)$ and $\overrightarrow{Tr}(p - a, q - b, r - c, s - d)$ their sum $\overrightarrow{Tr}(a, b, c, d) \boxplus \overrightarrow{Tr}(p - a, q - b, r - c, s - d)$ is determined as follows:

$$\begin{aligned} &\overrightarrow{Tr}(a, b, c, d) \boxplus \overrightarrow{Tr}(p - a, q - b, r - c, s - d) = \\ &= \begin{cases} \overrightarrow{Tr}(\min\{p, q\}, q, r, \max\{r, s\}), & (q < r) \vee (q = r \wedge p \leq s), \\ \overrightarrow{Tr}(\max\{p, q\}, q, r, \min\{r, s\}), & (q > r) \vee (q = r \wedge p > s). \end{cases} \end{aligned} \tag{5}$$

where the symbol $\boxplus$ denotes the addition operator.

For the case of any real number $\beta \in \mathbb{R}$ and any TrOFN $\overrightarrow{Tr}(a, b, c, d)$ their dot product can be calculated as follows:

$$\beta \odot \overrightarrow{Tr}(a, b, c, d) = \overrightarrow{Tr}(\beta \cdot a, \beta \cdot b, \beta \cdot c, \beta \cdot d) \tag{6}$$

where the symbol $\odot$ denotes the dot product operator.

3 Present value of portfolio

The present value (PV) is defined as a present equivalent of a cash flow in a given time in the present or future [6]. It is commonly accepted that the PV of a future cash flow can be imprecise[4]. The natural consequence of this approach is estimating PV with fuzzy numbers. A detailed description of the evolution of this particular model can be found in [7]. Among other things, an imprecise PV may be evaluated by TrOFN [4,10].

By a financial portfolio we will understand an arbitrary, finite set of financial assets. We consider a multi-assets portfolio π, consisting of assets Y_i $(i = 1,2, \dots, n)$. Each of these assets $Y_i \in \pi$ is characterized by monotonic sequence $\left(V_s^{(i)}, V_f^{(i)}, \check{C}^{(i)}, V_l^{(i)}, V_e^{(i)}\right) \subset \mathbb{R}^+$ where:

- $\left[V_s^{(i)}, V_e^{(i)}\right] \subset \mathbb{R}^+$ is interval of all possible PV' values,
- $\left[V_f^{(i)}, V_l^{(i)}\right] \subset \left[V_s^{(i)}, V_e^{(i)}\right]$ is interval of all prices which do not perceptible differ from market price $\check{C}^{(i)}$.
- $\check{C}^{(i)} \in \left[V_f^{(i)}, V_l^{(i)}\right]$ is market price of the asset Y_i.

Then we estimate the oriented PV (OPV) of any asset $Y_i \in \pi$ by TrOFN $\overrightarrow{PV_i} = \overrightarrow{Tr}\left(V_b^{(i)}, V_f^{(i)}, V_l^{(i)}, V_e^{(i)}\right)$. If we predict a rise in market price of the asset Y_i then its OPV is determined as positively oriented TrOFN. If we predict a fall in market price of the asset Y_i then its OPV is determined as negatively oriented TrOFN. In this way we distinguish the rising securities' portfolio $\pi^+ = \left\{Y_i \in \pi: \overrightarrow{PV_i} \in \mathbb{K}^+ \cup \mathbb{R}\right\}$ and the falling securities' portfolio $\pi^- =$

[4] The imprecision is understood as superposition of ambiguity and indistinctness of information. Ambiguity can be interpreted as a lack of a clear recommendation between one alternative among various others. Indistinctness is understood as a lack of explicit distinction between recommended and not recommended alternatives.

$\{Y_i \in \pi: \overleftarrow{PV_i} \in \mathbb{K}^-\}$. The portfolio PV is always equal to the sum of its components' PV. For OPV case, the addition should be modeled by sum $\boxplus$. In [9] it is shown that a result of multiple additions $\boxplus$ depends on summands' permutation. It implies that then portfolio's OPV given as any multiple sum of components' OPV is not explicitly determined. Therefore, in [10] some reasonable method of calculating portfolio's OPV is proposed.

At the first we calculate OPV of the rising securities' portfolio π^+

$$\overleftarrow{PV}^{(\pi+)} = \boxed{+}_{Y_i \in \pi^+} \overleftarrow{PV}^{(i)} \tag{7}$$[5]

and PV of the falling securities' portfolio π^-

$$\overleftarrow{PV}^{(\pi-)} = \boxed{+}_{Y_i \in \pi^-} \overleftarrow{PV}^{(i)}. \tag{8}$$

The both above OPVs are determined explicitly [9]. Moreover, single addition $\boxplus$ is commutative. Therefore, in [10] the OPV $\overleftarrow{PV}^{(\pi)}$ of portfolio π is explicitly determined as the sum

$$\overleftarrow{PV}^{(\pi)} = \overleftarrow{PV}^{(\pi+)} \boxplus \overleftarrow{PV}^{(\pi-)}. \tag{9}$$

4 Japanese candle as oriented fuzzy present value

Japanese candles [5] are very useful tool supporting investors' decision on exchange market. From the essence of the Japanese candles follows that they can be considered as an ambiguous PV estimation. In [3] it is shown that any Japanese candlestick can be represented by TrOFN $\overleftarrow{\mathcal{JC}} = \overleftarrow{Tr}(Pb, Po, Pc, Pf)$, where: Pb is back price, Po is open price, Pc is close price, Pf is face price. The back price Pb and the face price Pf are determined with using the extreme prices: the minimal price[6] Pl and the maximal price[7] Ph. All Japanese candles can be divided into three groups as follows:

- "white candle" $(Po < Pc)$ where the back price Pb is equal to the minimal price Pl and the face price Pf is equal to the maximal price Ph,
- "black candle" $(Po > Pc)$ where the back price Pb is equal to the maximal price Ph and the face price Pf is equal to the minimal price Pl,
- "Doji" $(Po = Pc)$ described by TrOFN with the orientation determined as direction from the earlier extreme price to the later extreme price.

Any Japanese candle may be used as OPV of asset unit. We consider a multi-assets portfolio π, consisting of securities' blocks B_i $(i = 1,2, \dots, n)$. The block B_i contains only β_i units of the security Y_i which is characterized by its Japanese candle $\overleftarrow{\mathcal{JC}}^{(i)}$. It implies that OPV of the block B_i is given as follows

$$\overleftarrow{PV}^{(i)} = \beta_i \odot \overleftarrow{\mathcal{JC}}^{(i)}. \tag{10}$$

As a result, together with (5), (6) and (7) we obtain that portfolio OPV can be calculated in the following way

$$\overleftarrow{PV}^{(\pi)} = \left(\boxed{+}_{\overleftarrow{\mathcal{JC}}^{(i)} \in \mathbb{K}^+} \beta_i \odot \overleftarrow{\mathcal{JC}}^{(i)}\right) \boxplus \left(\boxed{+}_{\overleftarrow{\mathcal{JC}}^{(i)} \in \mathbb{K}^-} \beta_i \odot \overleftarrow{\mathcal{JC}}^{(i)}\right). \tag{11}$$

Above OPV orientation is a prediction of further trend in global quotation of portfolio.

5 Case study

After closing the session on the Warsaw Stock Exchange on March 5, 2018, we evaluate the portfolio π of:

- the block B_1 of 10 shares of the ASSECOPOL (ACP),
- the block B_2 of 20 shares of the CYFROWY POLSAT (CPS),
- the block B_3 of 30 shares of the ENERGA (ENG),
- the block B_4 of 5 shares of the JSW (JSW),
- the block B_5 of 5 shares of the KGHM (KGH),
- the block B_6 of 10 shares of the LOTOS (LTS),

[5] The symbol $\boxed{+}$ denotes the multiple addition $\boxplus$

[6] In the original Japanese candlesticks' terminology minimal price is called low price.

[7] In the original Japanese candlesticks' terminology maximal price is called high price.

	Quotations			
Stock company	**Open price**	**Minimal price**	**Maximal price**	**Close price**
ACP	45.90	45.48	45.90	45.50
CPS	22.82	22.76**	22.92*	22.82
ENG	10.19	10.14	10.22	10.17
JSW	92.54	92.24*	92.80**	92.54
KGH	103.05	102.65	103.90	103.60
LTS	56.56	56.28	56.70	56.40
OPL	5.76	5.75	5.90	5.90
PGE	10.39	10.33	10.39	10.35
PKO	42.61	42.61	43.22	43.22
$*$ earlier extreme price; $**$ later extreme price				

Table 1. Observed quotations on the Warsaw Stock Exchange from 4 p.m. to 5 p.m. on March 15, 2018.

Source: [12]

- the block B_7 of 100 shares of the ORANGEPL (OPL),
- the block B_8 of 25 shares of the PKOBP (PGE),
- the block B_9 of 10 shares of the PKOBP (PKO).

In the Table 1 we present the stock quotations of these shares observed on the Warsaw Stock Exchange from 4 p.m. to 5 p.m. on March 5, 2018. We notice that:

- the stock companies KGH, OPL and PKO are evaluated by "white candles", which predicts a rise in market price;
- the stock companies ACP, ENG, LTS and PGE are evaluated by "black candles", which predicts a fall in market price;
- the stock companies CPS and JWS are evaluated by "Doji".

The detailed CPS and JSW quotations recorded in the analyzed time interval are presented in the Figure 1 and 2.

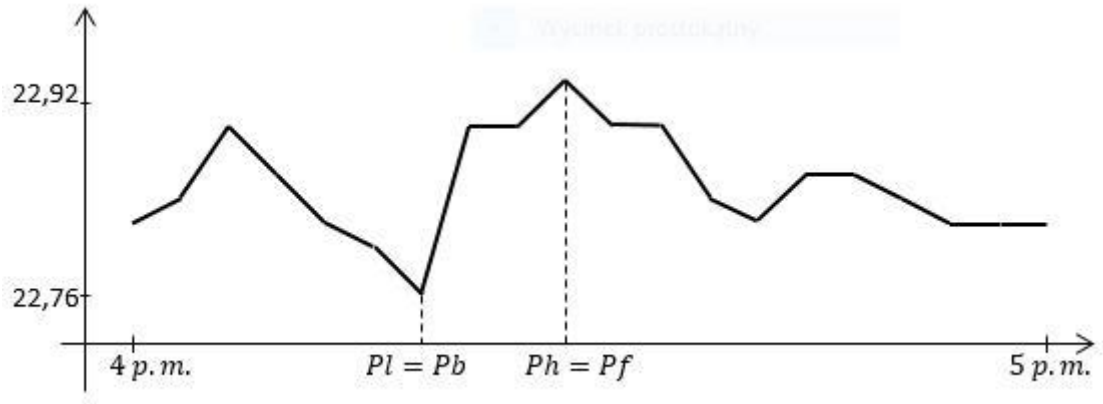


Figure 1. CPS quotations from 4 p.m. to 5 p.m. on March 15, 2018

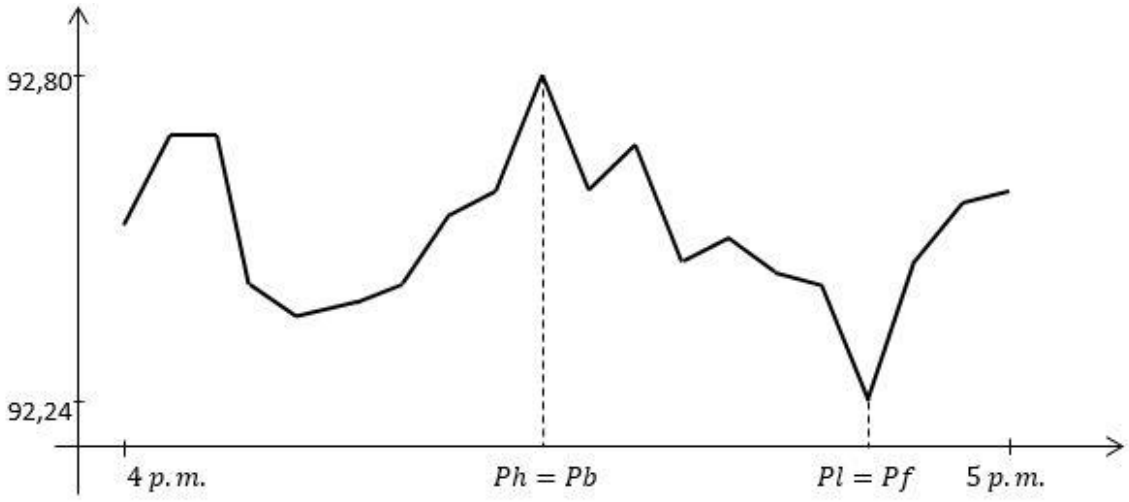


Figure 2. JSW quotations from 4 p.m. to 5 p.m. on March 15, 2018

We notice that:
- the stock company JWS is evaluated by positively oriented "Doji", which predicts a rice in market price,
- the stock company CPS is evaluated by negatively oriented "Doji", which predicts a fall in market price.

Let us evaluate the portfolio components with use their oriented fuzzy PV determined by Japanese candlesticks. If these Japanese candlesticks are represented by TrOFN then we have:

- $\overrightarrow{\mathcal{JC}}_{ACP} = \overrightarrow{\mathcal{Tr}}(45.90;\ 45.90;\ 45.50;\ 45.48)$,
- $\overrightarrow{\mathcal{JC}}_{CPS} = \overrightarrow{\mathcal{Tr}}(22.92;\ 22.82;\ 22.82;\ 22.76)$,
- $\overrightarrow{\mathcal{JC}}_{ENG} = \overrightarrow{\mathcal{Tr}}(10.22;\ 10.19;\ 10.17;\ 10.14)$,
- $\overrightarrow{\mathcal{JC}}_{JSW} = \overrightarrow{\mathcal{Tr}}(92.24;\ 92.54;\ 92.54;\ 92.80)$,
- $\overrightarrow{\mathcal{JC}}_{KGH} = \overrightarrow{\mathcal{Tr}}(102.65;\ 103.05;\ 103.60;\ 103.90)$,
- $\overrightarrow{\mathcal{JC}}_{LTS} = \overrightarrow{\mathcal{Tr}}(56.70;\ 56.56;\ 56.40;\ 56.28)$,
- $\overrightarrow{\mathcal{JC}}_{OPL} = \overrightarrow{\mathcal{Tr}}(5.75;\ 5.76;\ 5.90;\ 5.90)$,
- $\overrightarrow{\mathcal{JC}}_{PGE} = \overrightarrow{\mathcal{Tr}}(10.39;\ 10.39;\ 10.35;\ 10.33)$,
- $\overrightarrow{\mathcal{JC}}_{PKO} = \overrightarrow{\mathcal{Tr}}(42.61;\ 42.61;\ 43.22;\ 43.22)$.

In this way we distinguish the rising shares' portfolio $\pi^+ = \{JSW, KGH, OPL, PKO\}$ and the falling shares' portfolio $\pi^- = \{ACP, CPS, ENG, LTS, PGE\}$ as follows. Using the identity (8), for each considered block B_i $(i = 1,2, \dots, 9)$ of shares we calculate their oriented fuzzy $\overrightarrow{PV_i}$ as follows:

- $\overrightarrow{PV_1} = 10 \odot \overrightarrow{\mathcal{JC}}_{ACP} = \overrightarrow{\mathcal{Tr}}(459.00;\ 459.00;\ 455.00;\ 454.80)$,
- $\overrightarrow{PV_2} = 20 \odot \overrightarrow{\mathcal{JC}}_{CPS} = \overrightarrow{\mathcal{Tr}}(458.40;\ 456.40;\ 456.40;\ 455.20)$,
- $\overrightarrow{PV_3} = 30 \odot \overrightarrow{\mathcal{JC}}_{ENG} = \overrightarrow{\mathcal{Tr}}(306.60;\ 305.70;\ 305.10;\ 304.20)$,
- $\overrightarrow{PV_4} = 5 \odot \overrightarrow{\mathcal{JC}}_{JSW} = \overrightarrow{\mathcal{Tr}}(461.20;\ 462.70;\ 462.70;\ 464.00)$,
- $\overrightarrow{PV_5} = 5 \odot \overrightarrow{\mathcal{JC}}_{KGH} = \overrightarrow{\mathcal{Tr}}(513.25;\ 515.25;\ 518.00;\ 519.50)$,
- $\overrightarrow{PV_6} = 10 \odot \overrightarrow{\mathcal{JC}}_{LTS} = \overrightarrow{\mathcal{Tr}}(567.00;\ 565.60;\ 564.00;\ 562.80)$,
- $\overrightarrow{PV_7} = 100 \odot \overrightarrow{\mathcal{JC}}_{OPL} = \overrightarrow{\mathcal{Tr}}(575.00;\ 576.00;\ 590.00;\ 590.00)$,
- $\overrightarrow{PV_8} = 25 \odot \overrightarrow{\mathcal{JC}}_{PGE} = \overrightarrow{\mathcal{Tr}}(259.75;\ 259.75;\ 258.75;\ 258.25)$,
- $\overrightarrow{PV_9} = 10 \odot \overrightarrow{\mathcal{JC}}_{PKO} = \overrightarrow{\mathcal{Tr}}(426.10;\ 426.10;\ 432.20;\ 432.20)$.

At the first we calculate oriented fuzzy PVs of the rising shares' portfolio π^+ denoted by the symbol $\overleftarrow{PV}^{(\pi+)}$ and PV of the falling shares' portfolio π^- denoted by the symbol $\overleftarrow{PV}^{(\pi-)}$. Due (5) and (6) we obtain

$$\overleftarrow{PV}^{(\pi+)} = \overleftarrow{PV}_4 \boxplus \overleftarrow{PV}_5 \boxplus \overleftarrow{PV}_7 \boxplus \overleftarrow{PV}_9 = \overleftarrow{\mathcal{T}r}(1975.55; 1980.05;\ 2002.90;\ 2005.70),$$

$$\overleftarrow{PV}^{(\pi-)} = \overleftarrow{PV}_1 \boxplus \overleftarrow{PV}_2 \boxplus \overleftarrow{PV}_3 \boxplus \overleftarrow{PV}_6 \boxplus \overleftarrow{PV}_8 = \overleftarrow{\mathcal{T}r}(2050.75;\ 2046.45; 2039.25;\ 2035.25).$$

In the last step, due (7), we determine oriented fuzzy PV of portfolio π in as the revised sum

$$\overleftarrow{PV}^{(\pi)} = \overleftarrow{PV}^{(\pi+)} \boxplus \overleftarrow{PV}^{(\pi-)} = \overleftarrow{\mathcal{T}r}(4026.30;\ 4026.50;\ 4042.15;\ 4042.15).$$

We notice that the portfolio π is evaluated by negatively oriented OPV, which predicts a fall in its market value.

6 Summary

In this paper it is shown, that for any portfolio consisting of assets evaluated with oriented fuzzy PV estimated by TrOFN, the portfolio's PV is may be determined explicitly. The authors are convinced that the proposed method of PV addition is well-reasoned. On the other hand, in [9] it is shown that a result of multiple additions $\boxplus$ depends on summands' permutation. It implies that methods shown this paper is particular methods of portfolio PV determination. Obtained in this way PV can significantly differ in the order of addition oriented fuzzy PV of portfolio's components. This causes that the portfolio management methods with oriented fuzzy PV can significantly differ from the portfolio management methods with fuzzy PV. Never less, in [10] it is shown that the rising securities, portfolios π^+ and the failing securities π^- may be analyzed with use the results obtained in [11]. This is sufficient to manage portfolio risk, because of only the rising securities papers can get BUY or ACCUMULATE recommendations and only falling securities can get SELL or REDUCE recommendations.

On the other hand, obtained here results may be applied for calculating expected return rate (see [8]) and expected discount factor (see [4]) of portfolio with components evaluated by means of Japanese candles.

References

[1] Dubois D., Prade H. (1978) Operations on fuzzy numbers, *International Journal of System Science* 9, 613-629, doi:10.1080/00207727808941724

[2] Kosiński, W., Prokopowicz, P., Ślęzak D. (2002) Fuzzy numbers with algebraic operations: algorithmic approach. In M. Klopotek M., S.T.Wierzchoń, M. Michalewicz M., (Eds.) *Proc.IIS'2002*,(pp 311-320) Sopot, Poland,: Physica Verlag, Heidelberg.

[3] Łyczkowska-Hanćkowiak A., Piasecki K., (2018), On representation of Japanese candlesticks by ordered fuzzy numbers, In W. Szkutnik et al (Eds.) *9th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Winter Edition" Conference Proceedings,* (pp 61-68), Wyd. UE, Katowice 2018.

[4] Łyczkowska-Hanćkowiak A., Piasecki K., (2018),The expected discount factor determined for present value given as ordered fuzzy number, In W. Szkutnik et al (Eds.) *9th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Winter Edition" Conference Proceedings*, (pp 69-65), Wyd. UE, Katowice 2018.

[5] Nison S., (1991), *Japanese Candlestick Charting Techniques*, New York Institute of Finance, New York

[6] Piasecki K., (2012), Basis of financial arithmetic from the viewpoint of the utility theory, *Operations Research and Decisions*, 22(3), 37-53.

[7] Piasecki K. (2014), On Imprecise Investment Recommendations, *Studies in Logic, Rhetoric and Grammar*, 37, 179-194, doi: 10.2478/slrg-2014-0024.

[8] Piasecki K. (2017) Expected return rate determined as oriented fuzzy number, In P. Pražak, *35th International Conference Mathematical Methods in Economics Conference Proceedings*, (pp 561-565) Gaudeamus, University of Hradec Králové.

[9] Piasecki K., (2018) Revision of the Kosiński's theory of ordered fuzzy numbers, *Axioms* 7(1), 16, doi: 10.3390/axioms7010016

[10] Piasecki K., Łyczkowska-Hanćkowiak A., (2018) Present value of portfolio of assets with present values determined by trapezoidal ordered fuzzy numbers, *Operations Research and Decisions*, 28(2), (accepted for publication).

[11] Piasecki K., Siwek J., (2018) Two-assets portfolio with trapezoidal fuzzy present values, , In W. Szkutnik (Eds.) 9th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Winter Edition" Conference Proceedings, (pp.76-84) Wyd. UE, Katowice.

[12] https://www.bankier.pl/inwestowanie/profile/quote.html?symbol=WIG20 (accessed on 07.03.2018)

Exchange rate modeling with use of relative binary representation

Krzysztof Piasecki[1], Michał Dominik Stasiak[2]

Abstract. Many methods of exchange rate modeling are based on candlestick representation. This kind of representation causes a decrease in the value of information, resulting in a lower quality of modeling. Therefore, the binary representation seems to be more effective. In the binary representation, an increase of the price by assumed discretization unit is assigned a number ‚1', and the decrease is assigned ‚0'. In this paper we propose a new method of exchange rate trajectory binarization, that is, a relative binary representation, in which changeable discretization unit is directly proportional to the current exchange rate. Also, in the article, we performed a detailed comparative analysis of the possibilities of assessing the probability of future exchange rates' changes distribution for each considered binary representation. Research was performed with use of five-year quotations (2013-2018), for the AUD/NZD currency pair. During the research we used dedicated software, written in MQL4 and C++ languages.

Keywords: forex, exchange rate modeling, binary representations

JEL Classification: F31
AMS Classification: 90C15

1 Introduction

Exchange rates change very often. We can describe exchange rates process by its binary representation [2]. In the binary representation, an increase of the exchange rate by assumed value (discretization unit) is assigned a number ‚1', and the decrease is assigned ‚0'. In this paper we take into account the absolute binary representation determined by constant discretization unit and the relative binary representation determined by changeable discretization unit which is directly proportional to the current exchange rate.

The main goal to this paper is to compare representation utility for considered binary representation of exchange rates. Applied utility is defined as the Shannon's entropy [9] of the probability of future exchange rates' changes distribution. Research will be performed with use of five-year quotations (2013-2018), for the AUD/NZD currency pair.

2 Binary representations of an exchange rate

Exchange rates change with a very high frequency. The course change averagely each 1-2 seconds. Many of the changes, with a very small range of 1-2 pips, in fact have the character of random fluctuations (noise). Because of the reasons stated above, the tick data need to be formatted before further analysis. The domineering data representation in case of broker platforms (e.g. MetaTrader, JForex), is the candlestick representation. In this kind of representation, the trajectory changes are represented by four values: opening course, closing course, maximal and minimal price achieved in a given time period. This kind of course trajectory presentation is used in most of technical analysis methods: in methods of visual analysis, in appointing averages and indicators. This representation is also widely used in scientific research [1,3,6].

Candlestick representation, despite the vast application, causes a significant loss of informative value. The exchange rate changes on the currency market differ in their range. For example, during a publication of important macroeconomic data, the rate fluctuations are visibly higher than in case of night hours. In order to analyse changes with a range of, let us say, 20 pips, in case of candlestick representation one has to choose an appropriate timeframe. Even in case of the smallest of widely used time frames – equal to 1 minute – we achieve a candlestick representation as a series of registered candles with a range smaller for the case of the night hours. On the other hand, one candlestick in the period of high market trade can represent few or even few dozen changes of considered range, which will be ignored. The trajectory in a candlestick representation is also used in testing HFT strategies. When considering a system in which "take profit" (TP) and "stop loss" (SL) levels are placed e.g. 20 pips from

[1] Poznań University of Economics and Business, Department of Investment and Real Estate, Niepodległości 10, 61-875 Poznań, Poland e-mail: krzysztof.piasecki@ue.poznan.pl.
[2] 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.

the opening price, in many cases one cannot define which level was reached first, because both levels are included in a single candle. Because of this, in case of both analysis and strategy testing, application of candlestick representation can lead to false results, which stem from the loss of important information about the exchange rate trajectory variability [10].

2.1 Absolute binary representation

An alternative and more precise method of representation are the binary representations. The idea of binarization was first introduced in the visual analysis method, mainly point-symbolic method. The simplest binary representation, described in [10], is the absolute binary representation. The absolute binary representation consists in transforming the tick data into binary series. Absolute binarization algorithm indicates, for the initial course value $Q_{ask}(t')$ the change limit $Q_{ask}(t'')$ meeting the condition:

$$|Q_{ask}(t') - Q_{ask}(t'')| = \delta_{dys}, \tag{1}$$

where δ_{dys} is the discretization unit equal to given magnitude of unitary return, which is defined as a quotient of a return expressed in quoted currency by the amount of the base currency [7]. If the course drops below the lower limit, the binarization algorithm assigns a binary value ε = 0 to the change. In case of the price increase above the upper limit, the binarization algorithm assigns a binary value ε = 1 to the change. In ensuing steps, the algorithm appoints further binary values describing the changes in correspondence to the current rate value. In the Figure 1 an example of quotations' discretization is shown.

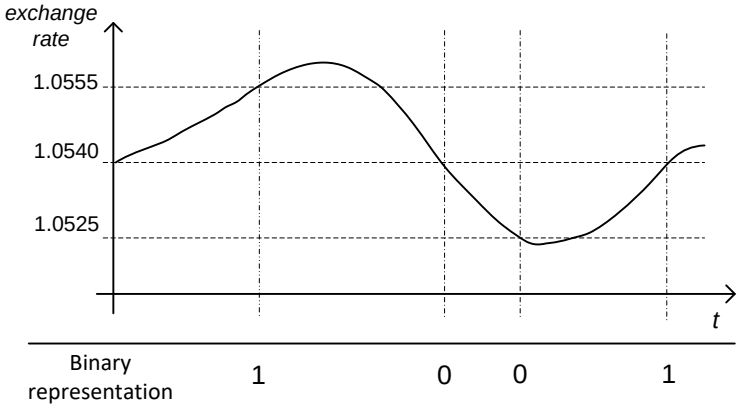


Figure 1 An exemplary conversion of tick data into binary data (δ_{dys} =15 pips)

On the currency market we can observe price gaps. This situation often has place e.g. in case of starting new quotations after weekends with the opening rate different than the closing rate on Friday. In these cases, the binarization algorithm checks, which value would have been reached and returns an appropriate result. Next binary value is then appointed in correspondence to the first price after the price gap.

In effect of binary algorithm operation, the process of exchange rate trajectory changes is presented as series of binary values $\{\varepsilon_i\}_{i=1}^{n}$, where n describes the number of registered changes. Therefore, each subseries $(\varepsilon_{\eta-m+1}, \varepsilon_{\eta-m+2}, \dots, \varepsilon_{\eta})$ is, in fact, a picture of most current history of changes in the exchange rate trajectory. Using so represented exchange rate one can use statistic methods to research possible relations on the currency market, e.g. wave relations [11, 12].

2.2 Relative binary representation

In the relative binary representation, we assume a constant in time value of magnitude of return rate. Assuming a constant value results in a change of X pips standing as a different percentage change, depending on the current price. To clarify, a change of 10 pips can stand as a change of 0,0009 of the course value, but after a half a year, it can stand as a 0,0015 of current course. Because of this, in the paper Authors propose and research a relative binary representation. Relative binarization algorithm indicates, for the initial course value $Q_{ask}(t')$ the change limit $Q_{ask}(t'')$ meeting the condition

$$\left|\frac{Q_{ask}(t') - Q_{ask}(t'')}{Q_{ask}(t')}\right| = r_{dys}, \tag{2}$$

where r_{dys} is the given magnitude of return rate. The actual form of the relative binary representation with the magnitude return rate described by (2) is, therefore, dependant on the assumed magnitude of the return rate. Let us consider the performance of discretization algorithm for an assumed discretization coefficient 0,001. Binarization algorithm appoints, for the initial rate value, the upper and lower change limit. It is equal to the positive and negative rate increase equal to 0,001 of the current rate. Next, analogously as in case of the absolute binary representation, if the course drops below the lower limit, the binarization algorithm assigns the binary value ε = 0 to the change. In case of an increase above the upper limit, the binarization algorithm assigns the binary value of ε = 1 to the change. In the following step, the algorithm appoints a limit based on the current exchange rate (according to formula (2)). In case of encountering a price gap, the algorithm of constructing a relative representation uses the same mechanisms as the algorithm for constructing the absolute binary representation.

3 Binary state model

Exchange rate given in a binary representation can be analysed in terms of possible regularities in investors' behaviours. In order to do so we can use state modelling. It assumes appointing market states that represent characteristic behaviours of the investors. The binary market state is defined as a subseries of binary values for an assumed time scope (*m* historical observations). The binary state is therefore an *m*-element permutation with repetitions from the set {0,1}:

$$\bar{\epsilon}_j = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m), \tag{3}$$

In this situation, the space of all *m*-element binary states ($\mathbb{E}_m$) is defined as a set of all *m*-element permutations with repetitions from the set {0,1}. The model allowing for describing and researching the dependencies between occurrence of behaviour schemas and the investors' reactions is the binary state model (BMS). The model describes binary market states. The binary state model $BMS(\mathcal{E}, m)$ uses two parameters: historical quotations in a binary representation ($\mathcal{E} = \{\varepsilon_i\}_{i=1}^{n}$) and the length of researched binary market states (*m*). Each state represents a particular schema of the investors' behaviour. In this situation, the space $\mathbb{E}_m$ of all *m*-character binary states can be defined as a set of all *m*-character permutations with repetition from the set {0, 1}. The state space $\mathbb{E}_m$ consists of exactly 2^m states. For example, in a BMS model, for which *m*=2, state (0,0) describes two rate drops of one discretization unit in a row, and state (0,1) describes a drop and an ensuing increase in the exchange rate, by the value of one discretization unit.

Binary transition function is a function $T_e: \mathbb{E}_m \longrightarrow \mathbb{E}_m \times \mathbb{E}_m$ described by the formula:

$$T_e(\bar{\epsilon}_j) = \{(\bar{\epsilon}_j, \bar{\epsilon}_k), (\bar{\epsilon}_j, \bar{\epsilon}_i)\}, \tag{4}$$

where

$$k = (2 \cdot j) \mathrm{mod} 2^m \tag{5}$$

$$l = k + 1. \tag{6}$$

Assuming the repeatability of investors' behaviours, we can appoint probabilities of transitions between so defined market states. So appointed probabilities can stand as premises in making investment decisions. In order to find these probabilities, for considered $BMS(\mathcal{E}, m)$ model, in the first step we appoint the probability distribution $p: \mathbb{E}_m \longrightarrow [0; 1]$ of each state occurrence. For given binary representation $\{\varepsilon_i\}_{i=1}^{n}$ and for the state $\bar{\epsilon}_j \in \mathbb{E}_m$ we have:

$$\mathrm{p}(\bar{\epsilon}_j) = \frac{n_j}{n - m + 1}, \tag{7}$$

where n_j is equal to the number of state $\bar{\epsilon}_j$ occurrences, and $n - m + 1$ is equal to the number of observations in general. Next, for each $\bar{\epsilon}_j \in \mathbb{E}_m$ we appoint the conditional probability distribution $p(\cdot \,|\bar{\epsilon}_j): T_e(\bar{\epsilon}_j) \longrightarrow [0; 1]$ of each transition occurrence:

$$p(\bar{\epsilon}_y | \bar{\epsilon}_j) = \frac{n_{j,y}}{n_j}, \tag{8}$$

where $n_{j,y}$ is equal to the number of occurrences of the transition between states $(\bar{\epsilon}_j, \bar{\epsilon}_y)$. In order to visualize the change model in the BMS model we can construct a transition graph, which shows the changes of exchange rate trajectory on the market. The detailed description of creating this kind of graph and its analysis is presented in [10]. In order to assess the informative value of the BMS model we calculate the expected Shannon's entropy measure [9], which is given by the formula:

$$E\big(BMS(\mathcal{E},m)\big) = -\frac{1}{ln2}\sum_{i=1}^{2^m}\sum_{(\bar{\epsilon}_j,\bar{\epsilon}_y)\in T_e(\bar{\epsilon}_j)} p(\bar{\epsilon}_j) * p(\bar{\epsilon}_y|\bar{\epsilon}_j) * \ln(p(\bar{\epsilon}_y|\bar{\epsilon}_j)), \tag{9}$$

The lower the entropy, the more information is included in the considered model and, in consequence, the better possible prognoses of future trajectory changes. Using BMS model allows for comparison between the quality of absolute and relative binary representation. If for the same time period and same tick data the entropy of BMS model for researched representation is lower, the higher the accuracy of applied quotation binarization mechanism.

4 The comparison of an absolute and relative course binarization

In order to compare the quality of researched representations Authors used tick data of AUD/NZD currency pair. This pair is the most popular currency pair among countries characterized by strong economic dependencies and located in the same region. These factors cause a similar influence of geopolitical situation on both economics, which allows for eliminating the influence of unexpected geopolitical factors on the exchange rate of the given currency pair. As a time scope of performed research Authors chose the five-year period 2013-2018). Many parameters of the currency market undergo significant changes: the character of trajectory changes, technology development (possibility of HFT systems programming), lowering of spreads and minimal deposits (which results in a bigger number of investors). Dependencies and dynamics, which could be observed when the currency market was limited to a narrow group of professional investors, differ from those observed nowadays. Because of these reasons, the analysis of older data would have been purposeless.

The Figure 2a presents dependencies between the BMS model entropy and the applied discretization unit in case of using the absolute binary representation. The Figure 2b shows also the number of observations 9the length of binary series). Analogous research was performed for BMS model with use of relative binary representation, in which he discretization unit is appointed according to the formula 2.

Based on presented results, for the magnitude of unitary return equal to 1 pip the entropy is higher than in case of 2 or 3 pip. This fact can be explained by the existence of the noise phenomenon on the currency market [4,5,7] (frequent and random small oscillations around the actual price) stemming from the fact that the magnitude rate in this situation is lower than the provision taken by the dealer. Analogous result was obtained in case of applying the relative binary representation (for the smallest magnitude of the return rate equal to 0,0001). Independently on the type of used representation, the binarization acts as a filter eliminating the noise. It is worth to note that the noise elimination causes a significant decrease of registered changes. Comparing the research results we can observe that for most cases, the application of the absolute discretization unit allows for just a slightly better quality of modeling. This effect comes from the fact that on the currency market most of the investors and the algorithm of automated trade base their decisions (e.g. indicating the TP and SL levels) on absolute values expressed in pips rather than percentage dependencies on the opening price. Simultaneously, it is valuable to note that the differences in the entropy values are not very high, and application of different binarization function in the absolute representation can possibly allow for obtaining better results, which is currently researched by the Authors.

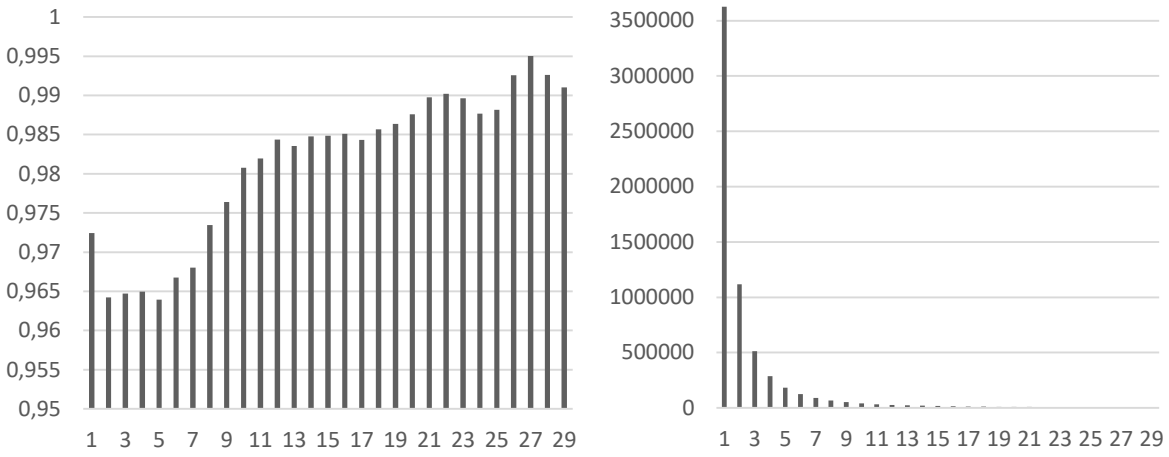


Figure 2 Entropy of BMS model (a) and the number of changes (b) in relations to the discretization unit in the absolute binary representation

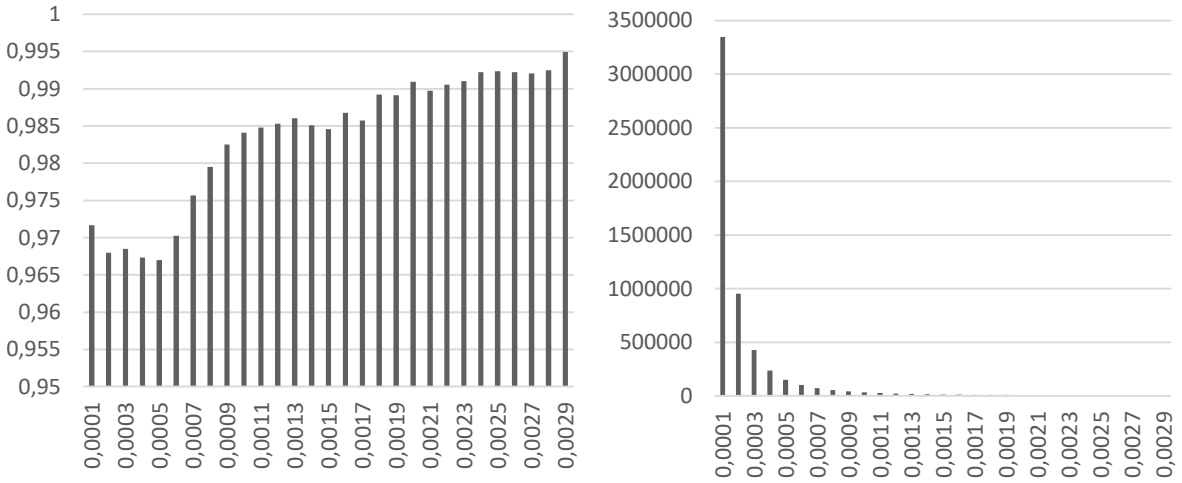


Figure 3 Entropy of BMS model (a) and the number of changes (b) in relations to the discretization unit in the relative binary representation

5 Summary

In the article, a new approach to the construction of binary representation is given. It assumes using a discretization function, which condition the current discretization level on the current market situation. The conditioning of binarization process on the current market situation opens new possibilities of modelling. In the paper Authors also propose a schema of representation quality assessment, based on entropy analysis in the state modelling.

In the article, an exemplary analysis of two representations was performed: absolute representation which use the proposed dependencies (constant percentage of current rate) and relative one, for a five-year quotation period of AUD/NZD currency pair. Both representations allow for noise filtration. Obtained results indicate a slightly higher quality of the absolute representation. The analysis of relative binary representation with proposed function (as a constant part of the current price) does not cause any improvement as compared to the absolute analysis of exchange rate. In this situation it is crucial to research if this kind of effect can be observed on the currency market for other currency pairs.

References

[1] Burgess, G. (2010) *Trading and Investing in the Forex Markets Using Chart Techniques* , Vol. 543, John Wiley and Sons.

[2] De Villiers, V. (1933) *The Point and Figure Method of Anticipating Stock Price Movements: Complete Theory and Practice*, Stock Market Publications.
[3] Kirkpatrick, C. D., and Dahlquist, J. A. (2010) *Technical analysis: the complete resource for financial market technicians*. FT press.
[4] Lo A. W., Mamaysky H., Wang J. (2000) Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation, *The Journal of Finance*, 55(4), pp. 1705-1770.
[5] Logue, D. E., and Sweeney, R. J. (1977) White noise in imperfect markets: the case of the franc/dollar exchange rates, *The Journal of Finance*, 32(3), 761-768.
[6] Murphy, J. J. (1986) *Technical analysis of the futures markets: A comprehensive guide to trading methods and applications*, Prentice Hall Press.
[7] Piasecki K., Stasiak M.D., (2018), The currency trading system with constant magnitude of unitary return, In W. Szkutnik et al (Eds.) *10th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Summer Edition" Conference Proceedings,* (pp 40-48), Publishing House of the University of Economics in Katowice
[8] Schlossberg, B. (2006) *Technical Analysis of the Currency Market*, John Wiley & Sons Inc.
[9] Shannon C. E. (1948) A Mathematical Theory of Communication, *Bell System Technical Journal*, Vol. 27.
[10] Stasiak M. D. (2017) Modelling of Currency Exchange Rates Using a Binary-Wave Representation, In Z. Wilimowska, L. Borzemski & J. Świątek (Eds.), *Proceedings of 38th In International Conference on Information Systems Architecture and Technology – ISAT 2017* (pp. 28-38), Springer.
[11] Stasiak M. D. (2017) Wave relations of exchange rates in binary-temporal representation, In Pavel Prazak , *Proceedings of 35th International Conference Mathematical Methods in Economics* (pp. 720-726), Gaudeamus.
[12] Stasiak, M. D. (2016) Modelling of currency exchange rates using a binary representation, In Z. Wilimowska, L. Borzemski, A. Grzech, J. Świątek (Eds.), *Proceedings of 37th International Conference on Information Systems Architecture and Technology – ISAT 2016* (pp. 153-163), Springer.

Decision analysis and multi-criteria evaluation using complex DEA models

Michal Pieter[1]

Abstract. Data envelopment analysis (DEA) is popularly known and frequently used as an evaluation tool. That said, in conjunction with various methods from Multi-criteria decision analysis (MCDA), it is often used to solve decision making problems as well. While examples of such combination are numerous, the cases where a particular DEA model is reinterpreted and reimplemented on its own to be used in decision making are much rarer. The last paper in the series provided some theoretical background, as well as issues faced, and examined under which conditions basic DEA models could be used in this way. This paper follows with an examination of some of the more complex DEA models and their applications in decision analysis, especially multi-criteria evaluation (MCE) of discrete variables. A particular focus is given on the branches of data envelopment analysis incorporating more real-world features, such as imprecise (fuzzy) data (f-DEA) or internal network structures (NDEA), in their models.

Keywords: Data envelopment analysis, DEA, MCDA, decision making, multi-criteria evaluation, MCE, fuzzy, network.

JEL Classification: C61, C67
AMS Classification: 90-02, 90-06, 90B50, 90C08, 90C29, 90C70

1 Introduction

Data envelopment analysis (DEA) is a branch of Operations research that allows one to look at a set of units (DMUs) and examine their relative efficiency. These units are all characterized by a certain number of inputs, which the unit consumes, and outputs, in turn produced by the unit. Units are assumed to be homogenous, with inputs and outputs (together factors) common to all of them. Today, as in the past, DEA is used in a wide variety of applications – public sector, private companies, health, education, environment protection, etc. Since it was first introduced by Charnes and Cooper [2] in 1978, it has radiated into a broad field experiencing vigorous growth in the decades since.

One of these areas is the exploration of the complex relationship between DEA and that of Multi-criteria decision making (MCDM, also analysis, MCDA). Similarities between these two fields have been noticed right from the beginning and authors have described them at times as being as both competing and complementary approaches. Belton and Stewart [1] provide a brief explanation of the issue. There have been many papers put forward proposing to combine DEA models and MCDA methods, in their various combinations, to solve a particular problem. Some opt to cut out the middleman altogether and use just DEA to solve decision problems.

The nature of the relationship between both fields and difficulties one runs into when attempting to use DEA models in and of themselves as an MCDA tool have been discussed in the previous paper of the series [9]. At the time of writing it is unavailable in English, but there are plans to publish the translated version as well. Brief theoretical and practical comparison between DEA and MCDA is to be found therein, which is why that topic is omitted from this follow-up. With the focus on Multi-criteria evaluation, concerning itself with discrete variables, some ways in which basic DEA models could be used to elicit a choice have also been examined previously. This paper follows directly from there, expanding its scope to some of the more complex models. Two broad categories of interest have been identified, first of which is fuzzy DEA (f-DEA), found in chapter 2, introducing the element of uncertainty and impreciseness into the models. Chapter 3 is then devoted to Network DEA (NDEA), with its notion of simple input-consuming, output-producing units extended to more realistic structures. Finally, some concluding remarks and proposals for further research are offered in the last chapter of this article.

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill sq. 4, 130 67 Praha 3, Czech Republic, michal.pieter@vse.cz.

1.1 Terminology

When discussing DEA and MCDA at the same time, some confusion may arise because of differing ways in which both fields use particular terms. Before proceeding, it may therefore be useful to clear up some terminology. MCDA, or more specifically MCE (multi-criteria evaluation), operates on a set of entities, most commonly referred to as variables, variants or options. Its goal is to select, or choose, a subset of these options, ideally one of them, as most preferred. To what measure each option is preferred depends on a number of criteria, each of which can prefer either lower or higher values. These values are the data and each represents the score of suitability of a given option, according to a particular criterion. DEA also deals with a certain set of entities, now called decision making units, or DMUs and units for short. DEA is traditionally used to evaluate all these entities, as opposed to selecting from them, and that evaluation depends on the values of inputs that a given unit consumes and outputs that it produces. The lower the values of inputs, or the higher those of outputs, the better. Inputs and outputs are also collectively called factors. It is clear from all of this that options in MCDA are analogous with units in DEA and criteria are analogous to factors, with inputs being minimizing (preferring lower values) and outputs maximizing (preferring higher). One can find further discussion comparing both fields in the previous paper [9] in this series.

Due to the nature of this article, the corresponding terms from DEA and MCDA will be used interchangeably throughout, depending on which one best serves in a particular context. Units, DMUs and options all denote the same thing. The term variables in this context will not be used, as that has another, obvious, meaning in DEA and which does not occur in the same fashion in MCE. Similarly, criteria and factors also will denote the same thing. In cases where a particular term may be ambiguous and its use is not obvious from context, clarification may be included, or two equal terms from both fields may be used in conjunction, to make clear what is being referred to.

2 Using fuzzy DEA models

In data envelopment analysis, the default assumption is that all the required data is known precisely. In truth, no measurement is ever 100% accurate and, outside of educational examples, all real data is imprecise to some degree. In most cases the simplification of ignoring this reality and using so-called crisp values is justified, mainly by making data gathering, calculation and interpretation that much easier. Sometimes, even despite the added complexity, it may be advantageous to capture this imprecise nature of the real world in the model, and thus obtain results that are both more reliable and more relevant. To that end, original data is usually obtained in the form of categories, an interval, or a set of data points. These are then transformed into various forms of fuzzy numbers, each defined by a membership function that assigns a value from 0 to 1 to every number on the real line, as originally proposed by Zadeh [16].

2.1 Fuzzy DEA with assurance region

One of the common problems that DEA models have to deal with, is how to avoid awarding a single factor the maximum weight and completely ignoring (having zero weight) the others. This is clearly important, as without this provision, each DMU would select for itself the one factor in which it excels. At best, all DMUs would choose the same factor, which would make the problem into a simple numerical ordering, undermining the very notion of complex evaluation. At worst, each DMU would choose a different factor, which is far more probable, especially when trying to use it in a MCDA context with many criteria/factors and few units/options. This would make any meaningful comparison impossible. Without going into too much detail, there are several ways of dealing with such a scenario and introduction of an assurance region (AR) is one of them. Simply put, DEA/AR, as first proposed by Thompson [12], restricts the weights so that for each pair of inputs the ratio of their weights falls into a certain range, the same being true for each pair of outputs. To compute these ranges, the relative importance of factors needs to be obtained, either by consulting a panel of experts, or some other method.

Liu [7] proposed to extend a DEA/AR model by introducing fuzzy numbers and to use such a model in an MCDA-type selection problem. He starts with a traditional output-oriented DEA model, then inserts the linearized version of assurance region constraints. These take the form $L_{O_i}/U_{O_j} \leq u_i/u_j \leq U_{O_i}/L_{O_j}$ for each combination of outputs O_i and O_j, where u, L and U are their respective weights, together with lowest and highest assigned importance score. Similarly, the constrains for each combination of inputs are inserted as well. By then introducing fuzzy numbers (with convex membership functions), the resulting efficiency also becomes fuzzy. To find the membership function of the efficiency scores, a pair of models is constructed to find the lower and upper bounds for the efficiency score for a given α-cut. In principle, by solving these models for all values of α, the function can be described entirely. In practice, by solving for several values of α, an approximation is found. The models for both lower and upper bounds are given as two-level programs, which he then proceeds to linearize into single-level linear forms. Both Jahanshahloo et al. [4] and Zhou [17] later found a mistake in this step and proposed a correction to both lower and upper bound programs that should be considered.

The proposed model is utilized in ranking 12 manufacturing systems and selecting the best one. Selection is based on 6 criteria (2 inputs, 4 outputs), two of which have crisp values, the rest are given as triangular fuzzy values. Since crisp values can be thought of as degenerate fuzzy numbers as well, the model can still be applied. By solving for differing values of α the membership function of fuzzy efficiency score of each DMU is approximated, in the form of a series of lower and upper bounds for different α-cuts. For increasing values of α these ranges shrink progressively, until at α=1 they degenerate into a single value. While Liu uses 11 α-cuts, he suggests that 3-4 are enough to rank the options. To that end, he uses an existing approach to construct an index from the membership function approximation. This index is finally suitable for full ranking of the options and marking the best one as the preferred choice.

Major advantage of Liu's approach is the fact that efficiency for each DMU is given as a fuzzy number itself, or approximation thereof. Many f-DEA models attempt to elicit crisp efficiency values from fuzzy data, which may give a false sense of precision. The drawback is that to get fuzzy efficiency scores for each DMU, $2\times n\times a$ programs would need to be computed, fortunately all linear. Interestingly, this makes it more suitable for use as an MCDA tool, where generally there are few options, relative to the number of criteria. The inclusion of assurance region is a sort of double-edged sword – while undermining the non-parametric nature of DEA and requiring the cooperation of experts, the setting of weight ratios allows one to eliminate solutions with unreasonable weight combinations.

2.2 Fuzzy DEA with preference ratio ranking

Khalili-Damghani, Tavana and Santos-Arteaga [5] propose a fuzzy model to aid in evaluation scenarios, as is common in DEA, as well as selection problems, as in MCDA. The purpose in creating the model lay in a concrete practical application – assisting a bank (anonymized in the paper) in expanding into emerging markets, by evaluating and ultimately selecting the best option from a list of 24 countries world-wide, based on 19 criteria. In regards to criteria, they introduce both desirable inputs and undesirable outputs, which have opposing optimization goals to common inputs and outputs, respectively. The authors start by structuring their model into three consecutive parts – a pre-processing stage designed to reduce the size of data, main module consisting of a fuzzy DEA model and a ranking module utilizing preference ratios. The pre-processing stage is designed to improve differentiating power of the model by aggregating highly correlated criteria, thus reducing their number. Because the model can be run without it, it suffices to say that a fuzzified form of weighted sums is used, in an iterative fashion so that the effect on final preferences is minimized. For further discussion, all of that can be safely ignored.

Main module consists of the fuzzy DEA model itself, which is based on a crisp, input-oriented BCC model, expanded to include desirable outputs and undesirable inputs. Resulting in a nonlinear model, a simple transformation is performed on these factors' values, returning to a linear form. Final crisp model is minimizing, so that a DMU with score 0 is efficient. In fuzzifying this crisp model, they consider trapezoidal fuzzy numbers, whose membership function takes in four parameters. Considering now this fuzzified model, for a given α-cut, it can be split in two and these in turn used to calculate the lower and upper efficiency bounds for a given DMU. Repeating this process for various α-cuts, the membership functions of efficiency scores themselves can be constructed – the larger the number of cuts, the more precise the approximation. This can however be a time-intensive exercise, so the authors propose an alternate approach. They use a clever substitution that incorporates, in a manner of speaking, the α-cut level into the model as a variable. This substitution relies on a linear shape of trapezoidal membership function, so one can imagine a similar model could be made (or rather simplified) for triangular fuzzy numbers, which are in effect their degenerate forms. After one further substitution, what one is left with are crisp and linear versions of the original lower and upper bound-calculating models, which are much easier to solve. Furthermore, as the α value is integrated into the models themselves, they need to be run just once each, automatically using the most favorable α-cut for each bound and unit.

Once the lower and upper bound efficiencies for each DMU are obtained, calculation proceeds to the ranking stage. The authors treat the efficiency bounds as a special case of trapezoidal fuzzy number (one could say rectangular), thus getting n (number of units) rectangles of height 1, as it were, with differing widths, sitting on a real number line. In the case where the rectangles do not overlap, ranking is trivial. If they do, however, they use what is called a preference function, which for a particular fuzzy number with membership function $\mu(x)$ and a given real value α returns, in essence, the ratio of area under $\mu(x)$, $x\geq\alpha$ to the total area under $\mu(x)$. Now, at every point α between the lower bound of the worst performing unit and upper bound of the best performer – lets call this section of the real line Ω – there is a unit which has the best value of preference function for that α. The set of points where that holds true for a given unit i is denoted as Ω_i, where $\Omega i \subset \Omega$. Preference ratio of each unit i is then defined simply as $|\Omega_i|/|\Omega|$, or the ratio of length of that section of Ω which prefers unit i to the length of entire Ω. The better the unit's ratio, the higher the ranking of that unit. Preference ratios for each DMU can be found algorithmically, and thus allow full ranking of all DMUs.

The major advantage of the proposed model is its relative simplicity in regards to required computational resources. The final model is not only linear, but also dispenses away with numerous α-cuts, that most fuzzy DEA models need to get a precise result. For n units only $2\times n$ executions would be needed, to get both lower and upper efficiency bounds for each unit. This is in contrast to Liu's model, as seen in the previous chapter. In addition to evaluation, the full ranking of units using preference ratios allows one to use it in MCDA scenarios as well, which the authors proceed to do with selecting the most suitable country for the bank to expand into.

2.3 Other approaches

There are several other approaches that attempt to use fuzzy DEA as a tool for option selection. Wang and Yan [14] use an assurance region model similar to that of Liu, discussed in chapter 2.1, to rank and select between different manufacturing models. Wang and Wei [13], while not strictly using fuzzy numbers, model uncertainty in data in a similar fashion. Assuming that real values lie within a region of uncertainty around some nominal values, they treat them as random perturbations. They propose several CCR-based models that allow full ranking under different scenarios, such as uncertainty only in inputs, in outputs, or both or different kinds of uncertainty.

3 Using network DEA models

While the previous chapter focused on better approximating the real world via recognizing and implementing the imprecise nature of data in the model, there are other ways of getting closer to reality. One of them is to take a closer look at the unit and to abandon a common simplification regarding their structure – namely that it has none. In traditional models all that is known about the unit is that it takes in some inputs and turns them into outputs. The nature of this transformation is assumed to be irrelevant and therefore any inner structure, or sub-processes going on inside the unit, are ignored. The unit is modelled as a black-box. As is the case with fuzzy data, this simplification yields simpler models, lessens the effort for both the analyst and the computer, and is a close enough approximation for most cases. When this proves to be insufficient, Network DEA allows one to peer inside the unit and model it as consisting of several sub-units (also sub-processes) arranged in a network structure, with each sub-process either connected to the outside world or tied to other sub-processes through common factors. This allows one not only to compute efficiencies of units more precisely and to better differentiate between them, but to also examine the efficiency of each individual sub-process.

3.1 Network-based approach

One of the problems in applying DEA directly to a decision problem is that often several DMUs are calculated to be efficient. The chief way of overcoming this is to modify the model so that full raking of DMUs is produced, from best to worst. While many such models exist for traditional DEA models (as examined in author's bachelor thesis [8]), when dealing with network structures, the literature is sparser. Liu and Lu [6] offer an approach that utilizes a network-based post-processing technique (not to be confused with the network structure itself) to produce such a ranking. They base their model on a simple 2-stage series network model where all outputs from stage (sub-process) 1 are immediately consumed as inputs of stage 2. With variable returns to scale, the chosen model tends to mark more DMUs as efficient (and never fewer) when compared with constant returns to scale. This model takes the form of a multi-objective program where the efficiencies of both stages are optimized.

In traditional DEA models there is a concept of reference units – efficient units that serve as models for the inefficient ones on how to structure their factors, were they to become efficient themselves. They can be identified by studying λ values, a common symbol denoting the weight of a given reference unit. Liu and Lu use these in order to rank the efficient units – the more often a given efficient unit serves as a reference to other units, and the more strongly, the better its final ranking. To further the discriminating power, they suggest to calculate the model multiple times, each time including some factors and excluding the others, in a manner postulated by Cinca and Molinero [10]. For best precision, this would entail running the model $n(2^m-1)(2^D-1)(2^s-1)$ times, where n is the number of DMUs and m, D, and s are numbers of inputs, intermediate factors, and outputs, respectively.

When the model is run for each DMU and for a given choice of factor combinations, the post-processing phase follows. While the computed efficiencies of both stages (sub-processes) for each unit and factor combination are discarded, the reference weights are used in further analysis. After normalization, designed to remove the effect of differing unit sizes, and averaging across factors, one is left with two sets of 0 to 1 numbers, $IOW1^t_{jk}$ and $IOW2^t_{jk}$. They represent a combined reference weight of all inputs and outputs of a particular unit k with regard to a reference unit j, for the first and second stage, respectively. Summing for all chosen combinations of factors t, one gets values of $A1_{jk}$ and $A2_{jk}$, which can be organized into two square matrices, $\mathbf{A1}$ and $\mathbf{A2}$, of size n. Finally, these matrices are key to eliciting all manner of information – for one, they allow to discriminate between efficient DMUs based on either eigenvectors or a so-called alpha-centrality score. Furthermore, one can rank units on how

well they serve as a benchmark for each individual factor. The model also provides further information, which is however not all that important in the context of using the model as an MCDA tool.

Main advantage of the proposed model lies in its ability to discriminate between efficient units and thus fully rank all the DMUs. To that end, it allows the analyst several options on how to rank them. They can focus more on one of the stages, or even incorporate the benchmarking strength of units with regard to a particular factor. One interesting further approach would be to take into account information already available in most MCDA scenarios – the weights of criteria. While diverging from DEA, where units set their own, most favorable weights, it might prove useful in aggregating the multiple rankings the proposed model generates. One could envision using the combined weights of input and output criteria-factors to produce an average ranking from stage 1 and stage 2 rankings, respectively. In a similar fashion, benchmark rankings for individual factors might be aggregated. The feasibility and practical applications of these aggregation approaches may warrant further examination.

There are two obvious disadvantages of the proposed model. First is the need to execute the mathematical program many times to obtain results for all, or most, possible combinations of included factors. The number of executions, given as $n(2^m-1)(2^D-1)(2^s-1)$, rises linearly with number of DMUs n, but exponentially with number of factors, m, D and s. This may prove especially troubling with MCDA applications, which tend to have have more criteria-factors, compared to units. In practical applications, one may expect hundreds or thousands of executions. The silver lining lies in the fact that the mathematical program is altogether linear, hence a single execution is usually quite fast. The second issue lies with extensive post-processing of data once all the programs have run. Both of these disadvantages seem to limit the usefulness of the proposed model when applied to practical MCDA problems and using traditional tools (spreadsheets and linear programming solvers). The proposed model would benefit greatly from a dedicated software tool, that could automatize many of the necessary tasks.

3.2 Other approaches

Few other authors attempt to model decision problems with network DEA. One of them are Yousefi, Shabanpour and Farzipoor Saen [15], who tried to apply it to a supply chain selection problem. To do so, they implemented some features of goal programming, a subset of multi-objective programming, into a series-structured network model. In DEA, each DMU chooses its own most suitable reference units that give it a set of reference factor values to aspire to, if it's not efficient. In goal programming, these reference factor values, or goals, are set beforehand by a manager and are common to all DMUs. The authors take a middle ground, where both approaches are used at the same time, their relative importance defined by a parameter elicited from the data itself. While their model still calls for setting the common goals by hand (although the best value for each factor could conceivably be used) it produces a full ranking of units. In fact, one can use for choice selection in an MCDA fashion, at least in principle any one of the network DEA models that allow the full ranking of DMUs, provided of course that criteria can be sensibly modelled using network structure. Models such as these include for example [3] or [11], among others. In cases where sub-process efficiencies cannot be combined to get overall efficiency and only single stage ranking is possible, a similar approach as the one discussed in the previous chapter could be used.

4 Concluding remarks

As mentioned in the introduction, this article is second in a series of papers aimed at illuminating the relationship between DEA and MCDA, particularly MCE. Their aim is to examine a variety of ways in which various DEA models can be used to tackle decision problems, be that alone or in conjunction with other MCDA methods. All of that contributes to an overarching goal of using this information in selecting the most appropriate approach in an application study, which should conclude the series and is planned to be released with other collaborators. This paper is an important step towards that goal, having identified several promising models that could be used.

The focus of the planned application paper revolves around provider selection in energy markets. A fuzzy DEA may prove useful in comparison with other fuzzy MCDA approaches. Based on the literature reviewed for this paper, the preference ratio model of Khalili-Damghani, Tavana and Santos-Arteaga, discussed in chapter 2.2 seems suitable enough. While its use of trapezoidal fuzzy number will probably prove to be unnecessary, it can also deal with, and be simplified for use with triangular fuzzy numbers as well. The modelling of provider selection as a network-structured process also remains a distant possibility. In that case Liu's and Lu's model described in chapter 3.1 seems like an intriguing option to examine in that particular MCDA setting, despite its computational and post-processing requirements. Examining how to include criteria weights may also prove interesting.

As the planned release of the concluding paper is still some time away, one or two papers continuing in this line of research may yet follow. Certainly, the research into joint use of various DEA models and MCDA methods has been very active for many years now and appears to continue to be so. There are therefore still many areas which may be explored and so further publications may follow even after this series is concluded.

Acknowledgements

This work was supported by grant no. IGA F4/20/2018 of the Internal Grant Agency of the Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] 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.

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

[3] Huang, J., Chen, J., and Yin, Z.: A Network DEA Model with Super Efficiency and Undesirable Outputs: An Application to Bank Efficiency in China. *Mathematical Problems in Engineering* (2014).

[4] Jahanshahloo, G.R., Sanei, M., Rostamy-Malkhalifeh, M., and Saleh, H.: A comment on "A fuzzy DEA/AR approach to the selection of flexible manufacturing systems." *Computers & Industrial Engineering* **56** (2009), 1713–1714.

[5] Khalili-Damghani, K., Tavana, M., and Santos-Arteaga, F.J.: A comprehensive fuzzy DEA model for emerging market assessment and selection decisions. *Applied Soft Computing* **38** (2016), 676–702.

[6] Liu, J.S., and Lu, W.-M.: Network-based method for ranking of efficient units in two-stage DEA models. *Journal of the Operational Research Society* **63** (2012), 1153–1164.

[7] Liu, S.-T.: A fuzzy DEA/AR approach to the selection of flexible manufacturing systems. *Computers & Industrial Engineering* **54** (2008), 66–76.

[8] Pieter, M.: Models for ranking decision making units in data envelopment analysis, Bachelor thesis, University of Economics, Prague, 2014.

[9] Pieter, M.: DEA as a multi-criteria evaluation tool. In: *Nové Trendy v Ekonometrii a Operačním Výzkumu [CD-ROM, Online]*, *Vydavateľstvo Ekonóm*, Bratislava, 2017, 118–125.

[10] Serrano-Cinca, C., and Molinero, C.: Selecting DEA specifications and ranking units via PCA. *Journal of The Operational Research Society - J OPER RES SOC* **55** (2004), 521–528.

[11] Tavassoli, M., Farzipoor Saen, R., and Faramarzi, G.R.: Developing network data envelopment analysis model for supply chain performance measurement in the presence of zero data. *Expert Systems* **32** (2015), 381–391.

[12] Thompson, R.G., Singleton, F.D., Thrall, R.M., and Smith, B.A.: Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas. *Interfaces* **16** (1986), 35–49.

[13] Wang, K., and Wei, F.: Robust data envelopment analysis based MCDM with the consideration of uncertain data. *Journal of Systems Engineering and Electronics* **21** (2010), 981–989.

[14] Wang, Y.F., and Yan, H.S.: A Fuzzy DEA/AR Method for Manufacturing Mode Selection. *Advanced Materials Research* **694–697** (2013), 3618–3625.

[15] Yousefi, S., Shabanpour, H., and Farzipoor Saen, R.: Selecting the best supply chain by goal programming and network data envelopment analysis. *RAIRO - Operations Research* **49** (2015), 601–617.

[16] Zadeh, L.A.: Fuzzy sets. *Information and Control* **8** (1965), 338–353.

[17] Zhou, Z., Yang, W., Ma, C., and Liu, W.: A comment on "A comment on 'A fuzzy DEA/AR approach to the selection of flexible manufacturing systems"' and "A fuzzy DEA/AR approach to the selection of flexible manufacturing systems." *Computers & Industrial Engineering* **59** (2010), 1019–1021.

Cluster Analysis of Digital Economy in the Old European Union Countries

Michał Bernard Pietrzak[1], Bartosz Ziemkiewicz[2]

Abstract. Improving conditions for developing digital economy has been considered as a basic determinant of keeping international competiveness of Europe. It has been pointed as one of the main objectives of the Europe 2020 plan. In this document an objective of reducing disparities, also in this field, has been given as an important task for the EU. The disparities in using the potential of the digital economy between the old EU-15 economies and the new member states has been an effect of structural underdevelopment of Eastern and Central Europe. However, recent years have also shown significant differences in this field among the EU-15. The main purpose of the article is to analyse the similarities between the old EU countries in regard to the digital economy development. In the research the Ward's clustering method was applied. Data provided by Eurostat were used for the analysis in the years 2011 and 2017. The research confirmed important disparities between the EU-15 countries, which indicate that the traditionally lower developed Southern European countries are not able to use the potential of the digital economy phenomenon for reducing the distance to the European leaders.

Key words: cluster analysis**,** digital economy, Ward's method, Kruskal's test

JEL Classification: C38

AMS classification: 91C20

1 Introduction

Taking advantage of the digital economy development potential has been considered as a possibility to decrease traditional disparities between the European economies. It has been pointed as a way to improve competitiveness of the European economy as a whole, which was stressed in such documents as Europe 2020 plan. Last global financial crisis has shown not only significant structural differences between old EU-15 economies and new EU member states, but also disparities in the case of the EU-15 countries. As a result, the main purpose of the article is to analyse similarities between the EU-15 countries in regard to the digital economy development. In the analysis the Ward's clustering method and Kruskal's test were applied. Data provided by Eurostat were used for the research in the years 2011 and 2017.

The current analysis is built on the previous research papers of the authors, which were devoted to the issue of the digital economy development at a regional level in the Visegrad countries [9, 10; 7].

2 Cluster analysis

Multidimensional comparative analysis methods are used to assess complex phenomena, which variability can be described by means of many features (diagnostic variables) [12; 4; 16; 19; 8; 31; 20; 21; 22; 28; 1; 14; 17; 13]. In scientific research the method of linear ordering and cluster analysis are most often used. Economic studies based on the use of these methods have been widely described in the literature and they concern many aspects such as: a level of socio-economic development [3; 37; 25; 23; 27, 36], an effectiveness of labour markets [26; 29; 30], a level of innovative potential and competitiveness [5; 6; 24; 32; 11; 15; 39; 38], or quality of institutions [2; 40].

Cluster analysis belongs to classification methods and allows to distinguish, within the considered economic phenomenon, relatively homogeneous groups of economic objects (clusters). Economic objects assigned to particular clusters are characterized by a high level of similarity to each other due to the studied phenomenon. The similarity between objects is determined on the basis of a selected metric, where Euclidean distance is most often used in economic studies. The classification of economic objects is usually performed using the selected agglomeration method. The Ward's agglomeration method will be used in the article, due to its properties of generating many clusters with a small number of objects [34], which is often needed in economic research. In the Ward's method clusters are separated on the basis of the criterion of minimizing the sum of squares of distances between objects [33; 18; 30].

[1] Michał Bernard Pietrzak, Nicolaus Copernicus University in Toruń, Faculty of Economic Sciences and Management, Ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: michal.pietrzak@umk.pl

[2] Bartosz Ziemkiewicz, Nicolaus Copernicus University in Toruń, Department of Probability Theory and Stochastic Analysis, ul. Chopina 12/18, 87-100 Toruń, Poland, e-mail: bartek@mat.umk.pl

According to the objectives of the research, the cluster analysis was used to assess the similarities between economies in regard to the level of development of the digital economy. Ten diagnostic variables were selected for the description of selected aspects of the digital economy, for which statistical data is available in the Eurostat database. The set of diagnostic variables is presented in Table 1, where the symbol, name, description and unit of measurement were included. All the diagnostic variables can be considered as stimulants. The selection of the diagnostic variables is based on the suggestions of Eurostat in regard to the digital economy assessment at national level, which was supported by previous literature review of the authors [9; 35].

Variable	Name	Descripiton	Unit
x1	Households - level of internet access	Household internet connection: all type	Percentage of households
x2	Households - type of connection to the internet	Household internet connection type: broadband	Percentage of households
x3	Households - availability of computers	Households having access to, via one of its members, a computer	Percentage of households
x4	Individuals - mobile internet access	Individuals used a laptop, notebook, netbook or tablet computer to access the internet away from home or work	Percentage of individuals
x5	Individuals - frequency of computer us	Frequency of computer use: daily	Percentage of individuals
x6	Individuals - computer use	Last computer use: within last 12 months	Percentage of individuals
x7	Individuals - frequency of internet use	Frequency of internet access: once a week (including every day)	Percentage of individuals
x8	Internet purchases by individuals	Last online purchase: in the last 3 months	Percentage of individuals
x9	E-government activities of individuals via websites	Internet use: interaction with public authorities (last 12 months)	Percentage of individuals
x10	E-commerce purchases	Enterprises having purchased via computer mediated networks	Percentage of individuals

Table 1 The set of diagnostic variables

3 Empirical results

In the article the cluster analysis for assessment of the level of the digital economy development in the EU-15 countries was carried out. The analysis was done separately for the years 2011 and 2017. The main restriction in regard to the period of the research was the availability of data. The obtained results allowed to separate groups of countries (clusters), which are similar to each other from the perspective of the level of digital economy development. Therefore, it was possible to compare the clusters over time.

In the first step of the study, a standardization of variables was made on the basis of the mean and standard deviation. Then, the distance matrix was determined with application of the Euclidean distance. In the next step, three clusters of countries were established using the Ward‘s method. The obtained results of the Ward‘s method were presented on dendograms (Figure 1), and the established composition of countries into clusters was given in Table 2.

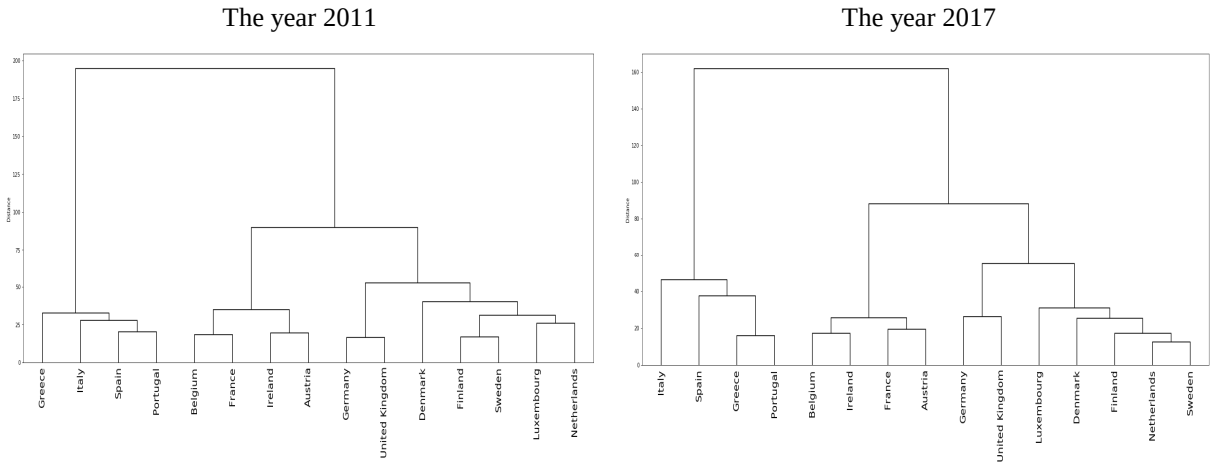


Figure 1 Dendrograms for the EU-15 in the years 2011 and 2017

2011		2017	
Cluster	**Country**	**Cluster**	**Country**
1	Greece	1	Greece
1	Italy	1	Italy
1	Spain	1	Spain
1	Portugal	1	Portugal
2	Belgium	2	Belgium
2	France	2	France
2	Ireland	2	Ireland
2	Austria	2	Austria
3	Germany	3	Germany
3	United Kingdom	3	United Kingdom
3	Denmark	3	Denmark
3	Finland	3	Finland
3	Sweden	3	Sweden
3	Luxembourg	3	Luxembourg
3	Netherlands	3	Netherlands

Table 2 Compositions of clusters in the years 2011 and 2017

In order to describe the obtained results, the average values of the diagnostic variables for each of the clusters were determined. The analysis of the mean values for the subsequent variables allowed to assess the level of digital economy development for countries from a selected cluster. Additionally, the Kruskal test was applied here, which allowed to check whether the determined mean values of variables differ significantly between each other for the obtained clusters (Table 3).

Based on the received values of Kruskal's statistics, significant differences were found between groups of countries assigned to individual clusters. In the cluster 1 one could find the countries with the lowest level of digital economy development. The cluster formed the countries of Southern Europe: Spain, Portugal, Italy and Greece. Such economies as: Austria, Belgium, Ireland, France have been assigned to cluster 2. These countries are characterized by an average level of the digital economy development. In the cluster 3 with the highest level of the digital economy development one could find: Denmark, Sweden, Netherlands, Sweden, Finland, United Kingdom, Luxembourg and Germany. It should be emphasized that countries from clusters 2 and 3 are characterized by a much better economic situation compared to the economies of the Southern Europe, which can be measured with basic standard macroeconomic indicators such as GDP per capita, GDP growth rate or level of government debt.

Variables								
Cluster	x1		x2		x3		x4	
	2011	2017	2011	2017	2011	2017	2011	2017
1	58.25%	77.25%	53.75%	76.25%	64.25%	73.25%	15.00%	29.25%
2	76.50%	87.25%	70.25%	84.75%	79.00%	84.50%	10.25%	20.50%
3	88.00%	95.43%	80.00%	94.00%	89.29%	93.86%	26.71%	40.29%
Kruskal statistics	12.14	12.17	10.63	11.70	12.17	12.39	7.06	8.74
p-value	0.002	0.002282	0.005	0.002876	0.002	0.002042	0.029	0.012642
Variables								
Cluster	x5		x6		x7		x8	
	2011	2017	2011	2017	2011	2017	2011	2017
1	46.00%	52.25%	61.25%	69.25%	52.50%	71.00%	13.00%	28.00%
2	63.00%	65.25%	80.75%	84.25%	74.75%	83.25%	35.00%	50.00%
3	76.57%	79.29%	90.43%	94.14%	85.43%	93.14%	54.00%	67.86%
Kruskal statistics	11.89	12.14	12.10	12.21	11.61	11.75	12.14	12.14
p-value	0.003	0.002	0.002	0.002	0.003	0.003	0.002	0.002

Variables				
Cluster	x9		x10	
	2011	2017	2011	2017
1	31.00%	42.25%	20.25%	27.75%
2	49.75%	60.00%	42.75%	51.50%
3	62.14%	73.14%	50.86%	53.86%
Kruskal statistics	9.51	8.16	8.37	7.69
p-value	0.009	0.017	0.015	0.021

Table 3 Average values of variables for the given clusters and Kruskal statistics for the year 2011 and 2017

4 Conclusions

The main objective of the research was to analyse similarities between the EU-15 economies in regard to the digital economy development. To obtain the aim of the article the Ward's clustering method was used for the study in the year 2011 and 2017. From the perspective of the dynamics of the phenomenon under research the given period should be considered as long enough for catching the potential changes in structural differences between the countries.

The conducted cluster analysis confirms the structural problems of the Southern European economies, which have been especially seen after last global financial crisis, with the real macroeconomic bankruptcy of Greece and long term stagnation of Spain, Italy and Portugal.

The most important empirical value added of the current research relates to the fact that the obtained results confirm high stability of disparities between the EU-15 countries. This obtained results compared with the experiences of the European economies after last global financial crisis brings significant long term growth policy implications. The study confirms that taking advantage of the digital economy for obtaining higher level of GDP growth in the case of countries at lower level of development (South European economies) is very difficult task, which can be reached only under conditions of fundamental structural macroeconomic reforms.

References

[1] Ahmed, R.R., Vveinhardt, J. & Streimikiene, D. (2017). Interactive digital media and impact of customer attitude and technology on brand awareness: evidence from the South Asian countries. *Journal of Business Economics and Management*, 18(6), 1115-1134.

[2] Balcerzak, A. P. (2009). Effectiveness of the Institutional System Related to the Potential of the Knowledge Based Economy. *Ekonomista*, 6, 711-739.

[3] Balcerzak, A. P. (2015). Europe 2020 Strategy and Structural Diversity Between Old and New Member States. Application of Zero Unitarization Method for Dynamic Analysis in the Years 2004-2013. *Economics & Sociology*, 8(2), 190-210.

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

[5] Balcerzak, A. P. (2016b). Technological Potential of European Economy. Proposition of Measurement with Application of Multiple Criteria Decision Analysis. *Montenegrin Journal of Economics*, 12(3), 7-17.

[6] Balcerzak, A. P. (2016c). Multiple-criteria Evaluation of Quality of Human Capital in the European Union Countries. *Economics & Sociology*, 9(2), 11-27.

[7] Balcerzak, A. P. (2017). Digital Economy in Czech Republic, Slovakia and Hungary. Measurement with TOPSIS Based on Entropy Measure for Objective Weighting. In T. Loster & T. Pavelka (Eds.). *The 11th International Days of Statistics and Economics. Conference Proceedings. September 8-10, 2016* (pp. 49-57). Prague: Libuse Macakova, Melandrium.

[8] Balcerzak, A. P., Kliestik, T., Streimikiene, D. & Smrčka L. (2017). Non-parametric approach to measuring the efficiency of banking sectors in European Union Countries. *Acta Polytechnica Hungarica*, 14(7), 51-70.

[9] Balcerzak, A. P. & Pietrzak, M. B. (2017a). Digital Economy in Visegrad Coutnries. Multiple-criteria decision analysis at Regional Level in the Years 2012 and 2015. *Journal of Competitiveness*, 9(2), 5-18.

[10] Balcerzak, A. P. & Pietrzak, M. B. (2017b). Digital economy in Polish regions. Proposal of measurement via TOPSIS with generalized distance measure GDM. In M. Papież and S. Śmiech (Eds.). *The 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*. *Conference Proceedings* (pp. 21-28). Cracow: Foundation of the Cracow University of Economics.

[11] Cheba, K. & Szopik-Depczyńska, K. (2017). Multidimensional comparative analysis of the competitive capacity of the European Union countries and geographical regions. *Oeconomia Copernicana*, *8*(4), 487-504.

[12] Erkut, B. (2016). Entrepreneurship and Economic Freedom: Do Objective and Subjective Data Reflect the Same Tendencies? *Entrepreneurial Business and Economics Review*, 4(3), 11-26.

[13] Grmanová, E. & Pukala, R. (2018). Efficiency of insurance companies in the Czech Republic and Poland. *Oeconomia Copernicana*, *9*(1), 71-85.

[14] Hajduk, S. (2018). Efficiency evaluation of urban transport using the DEA method. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 13(1), 141-157

[15] Kaynak, S., Altuntas, S. & Dereli, T. (2017) Comparing the innovation performance of EU candidate countries: an entropy-based TOPSIS approach. *Economic Research-Ekonomska Istraživanja*, 30(1), 31-54.

[16] Keshavarz Ghorabaee, M., Zavadskas, E.K., Amiri, M., & Antucheviciene, J. (2016). Evaluation by an Area-based Method of Ranking Interval Type-2 Fuzzy Sets (EAMRIT-2F) for Multi-criteria Group Decision-making. *Transformations in Business & Economics*, 15(3) (39), 76-95.

[17] Mačerinskienė, I. & Aleknavičiūtė, R. (2017). National intellectual capital influence on economic growth in the European Union countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(4), 573-592.

[18] Maknickiene, N., Lapinskaite, I. & Maknickas, A. (2018). Application of ensemble of recurrent neural networks for forecasting of stock market sentiments. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 13(1), 7-27.

[19] Mardani, A., Jusoh, A., Zavadskas, E.K., Kazemilari, M., Ungku, N.U.A. & Khalifah, Z. (2016). Application of Multiple Criteria Decision Making Techniques in Tourism and Hospitality Industry: a Systematic Review. *Transformations in Business & Economics*, 15(1)(37), 192-213.

[20] Meluzín, T., Balcerzak, A.P., Pietrzak, M. B., Zinecker, M., & Doubravský, K. (2018a). The impact of rumours related to political and macroeconomic uncertainty on IPO success: evidence from a qualitative model. *Transformations in Business & Economics*, 17, 2(44), 148-169.

[21] Meluzín, T., Pietrzak, M. B., Balcerzak, A. P., Zinecker, M., Doubravský, K., & Dohnal, M. (2017). Rumours Related to Political Instability and their Impact on IPOs: The Use of Qualitative Modeling with Incomplete Knowledge. *Polish Journal of Management Studies*, 16(2), 171-187.

[22] Meluzín, T., Zinecker, M., Balcerzak, A.P., Doubravský, K., Pietrzak, M. B., & Dohnal, M. (2018b), The timing of initial public offerings – non-numerical model based on qualitative trends. *Journal of Business Economics and Management*, 19(1), 63-79.

[23] Mikušová, M. (2017) To be or not to be a business responsible for sustainable development? Survey from small Czech businesses. *Economic Research-Ekonomska Istraživanja*, 30(1), 1318-1338.

[24] Monni, S., Palumbo, F., & Tvaronavičienė, M. (2017). Cluster performance: an attempt to evaluate the Lithuanian case. *Entrepreneurship and Sustainability Issues*, 5(1), 43-57.

[25] Novak, M. & Pahor, M. (2017). Using a multilevel modelling approach to explain the influence of economic development on the subjective well-being of individuals. *Economic Research-Ekonomska Istraživanja*, 30(1), 705-720.

[26] Pietrzak, M. B. (2016). The Problem of the Inclusion of Spatial Dependence Within the TOPSIS Method. *Montenegrin Journal of Economics*, 12(3), 69-86.

[27] Pietrzak, M. B. & Balcerzak, A. B. (2017). Economic development of Polish voivodeships in the years 2010-2014. Application of taxonomic measure of development with entropy weights. In M. Papież and S. Śmiech (Eds.), *The 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*. Cracow: Foundation of the Cracow University of Economics, 310-318.

[28] Pietrzak, M. B., Balcerzak, A. P., Gajdos, A. & Arendt, Ł (2017). Entrepreneurial environment at regional level: the case of Polish path towards sustainable socio-economic development. *Entrepreneurship and Sustainability Issues, 5*(2), 190-203.

[29] Pietrzak, M. B. & Ziemkiewicz, B. (2017). The use of random fields in the Modifiable Areal Unit Problem. In M. Papież and S. Śmiech (Eds.). *The 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings* (pp.319-328). Cracow: Foundation of the Cracow University of Economics.

[30] Rollnik-Sadowska, E. & Dąbrowska, E. (2018). Cluster analysis of effectiveness of labour market policy in the European Union. *Oeconomia Copernicana*, 9(1), 143-158.

[31] Shpolianska, I., Dolzhenko, A. & Glushenko, S. (2017). Fuzzy models of web services selection in the development of service-oriented information systems for small and medium enterprises. *Ekonomicko-manazerske spectrum*, 11(2), 64-75.

[32] Stankeviciene, J., Kraujaliene, L. & Vaiciukeviciute, A. (2017). Assessment of technology transfer office performance for value creation in higher education institutions. *Journal of Business Economics and Management*, *18*(6), 1063-1081.

[33] Trąpczyński, P., Jankowska, B., Dzikowska, M., & Gorynia, M. (2016). Identification of Linkages between the Competitive Potential and Competitive Position of SMEs Related to their Internationalization Patterns Shortly after the Economic Crisis. *Entrepreneurial Business and Economics Review*, 4(4), 29-50.

[34] Ward, J. H., Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58, 236–244.

[35] Wierzbicka, W. (2018). Information infrastructure as a pillar of the knowledge-based economy — an analysis of regional differentiation in Poland. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 13(1), 123-139.

[36] Vavrek R., Adamisin P. & Kotulic R. (2017). Multi-criteria Evaluation of Municipalities in Slovakia - case study in selcected districts. *Polish Journal of Management Studies*, 16(2), 290-301.

[37] Zavadskas, E. K., Govindan, K., Antucheviciene, J. & Turskis, Z. (2016). Hybrid multiple criteria decision-making methods: a review of applications for sustainability issues. *Economic Research-Ekonomska Istraživanja*, 29(1), 857-887.

[38] Zemlickienė, V., Bublienė, R. & Jakubavičius, A. (2018). A model for assessing the commercial potential of high technologies. *Oeconomia Copernicana*, 9(1), 29-54.

[39] Zygmunt, A. (2017). Innovation activities of Polish firms. Multivariate analysis of the moderate innovator countries. *Oeconomia Copernicana*, 8(4), 505-521.

[40] Żelazny, R., & Pietrucha, J. (2017). Measuring innovation and institution: the creative economy index. *Equilibrium. Quarterly Journal Of Economics And Economic Policy*, 12(1), 43-62.

Modeling E-Sports Matches Using Generalized Autoregressive Score Models

Mirek Pikhart [1], Vladimír Holý [2]

Abstract. In econometrics, a recent and popular way to model time series is the generalized autoregressive score model, also known as the dynamic conditional score model. It is an observation-driven model capturing dynamics of time-varying parameters by the autoregressive term and the scaled score of the conditional observation density. The GAS model can be based on any distribution. For example the GARCH model is a special case of the GAS model with normal distribution and time-varying variance while the ACD model is a special case with exponential distribution and time-varying mean.
We utilize the GAS methodology in modeling the results of e-sport professional competition, specifically the matches in Counter-Strike: Global Offensive video game. We model the results of matches using several methods, such as logistic regression or utilizing the Poisson distribution, comparing several approaches to the problem. Time-varying parameters represent the unobserved strength of individual teams.

Keywords: Sports Analytics, Count Time Series, Generalized Autoregressive Score Model.

JEL classification:
AMS classification:

1 Introduction

For a long time, the forecasting of match results has been highly popular, no matter what kind of sport it concerns. Originally starting with mostly static models, such as Reep and Benjamin's 1968 [15] paper proposing negative binomial distribution for modelling the number of goals scored and successful passes completed in a match. Other models built on similar foundations argued which distribution is the best fit, such as Maher [13], who proposed the Poisson distribution, or Fahrmeir and Tutz [6], using ordered logit models with random walks for forecasting both soccer and chess results.

Over time, the usage of dynamic models increased. Dixon and Coles [5] built on Maher's [13] foundation, improving on the bivariate Poisson model by realizing that the strengths of teams are not constant in time and including match results up to time t in their model when determining parameters for time t. Rue and Salvensen [16] went even further, proposing a completely dynamic team strengths in time, which allows for more accurate short-time predictions.

Koopman and Lit [9] proposed a Generalized Autoregressive Score (GAS) model to model the Premier League Soccer matches using the bivariate Poisson distribution and extended their study to include a Skellam distribution and ordered probit models in their 2017 paper [11]. Their models use a team's offensive and defensive strength as parameters together with a fixed home-side advantage, which reflects the fact that teams perform better in front of their home crowd.

GAS models are commonly used to analyse high-frequency economic data, such as Avdulaj and Kverav [1]'s paper on stock diversification or Zhao and van Wijnbergen [17], who modeled decision making in incomplete markets. Lucas, Schwaab and Zhang's [12] paper on sovereign credit risk of euro states showcases another common usage of GAS models, credit risk modeling.

We noticed that while there are dozens of articles on forecasting sports results, there are almost no papers published on the recently popularized eSports - 'a catchall term for games that resemble conventional sports insofar as they have superstars, playoffs, fans, uniforms, comebacks, and upsets. But all the action occurs online, and the contestants hardly move' as written by Jenny et al [8]. While searching for any papers concerning eSports statistics, a basic study on the effect of nationality on success in eSports [14] can be found together with some other articles concerning economics and other fields, but overall the statistics field is left completely unexplored.

[1]University of Economics, Prague, Winston Churchill Square 1938/4, 130 67 Prague 3, Czech Republic, xpikm05@vse.cz
[2]University of Economics, Prague, Winston Churchill Square 1938/4, 130 67 Prague 3, Czech Republic, vladimir.holy@vse.cz

In our article, we took the game Counter-Strike: Global Offensive, which is considered as one of the prime eSports titles. The game is played by two teams of five and can be covered by an adaptation of the soccer models as the two teams take turns playing offense and defense. The major difference is that the game takes place on a pre-selected map, which we presume has a significant effect on team performance.

2 Generalized Autoregressive Score Framework

Generalized autoregressive score (GAS) models [3, 4], also known as *dynamic conditional score* models [7] are time series models with time-varying parameters. The dynamics of time-varying parameters is captured by the autoregressive term and the score of the conditional observation density. In the case of unscaled score, the time-varying parameter f_t follow the recursion

$$f_{t+1} = C + Bf_t + A\nabla(x_t, f_t), \tag{1}$$

where C is the constant parameter, B is the autoregressive parameter, A is the score parameter and $\nabla(x_t, f_t)$ is the score given by

$$\nabla(x_t, f_t) = \frac{\partial \ln \mathrm{P}[X_t = x_t|f_t]}{\partial f_t}. \tag{2}$$

The GAS models belong to the class of observation-driven models according to [2]. Forecasting accuracy of the observation-driven models based on the score is comparable to the parameter-driven models [10]. An advantage of the observation-driven models including the GAS model is simple estimation by the maximum likelihood method.

3 Proposed Model

The proposed model is based on the GAS specification with Bernoulli distribution and time-varying probability of winning the match. This probability is transformed via logistic link to ensure its values fall within the $(0, 1)$ interval. This corresponds to the logistic regression model. The proposed model has the following structure.

If we consider the teams to have different strength on each map, then the probability of team i defeating team j on map m in time t is

$$p_{i,j,m,t} = \frac{1}{1 + e^{-(S_{i,t} + M_{i,m} - S_{j,t} - M_{j,m})}} \tag{3}$$

with the skill $S_{i,t}$ of team i following recursion

$$\begin{aligned}
S_{i,t+1} &= C_i + B_i S_{i,t} + A_i \frac{e^{S_{k,t}+M_{k,o}}}{e^{S_{i,t}+M_{i,o}} + e^{S_{k,t}+M_{k,o}}} && \text{if team } i \text{ defeated team } k \text{ on map } o \text{ in time } t, \\
S_{i,t+1} &= C_i + B_i S_{i,t} - A_i \frac{e^{S_{i,t}+M_{i,o}}}{e^{S_{i,t}+M_{i,o}} + e^{S_{k,t}+M_{k,o}}} && \text{if team } i \text{ lost to team } k \text{ on map } o \text{ in time } t, \\
S_{i,t+1} &= S_{i,t} && \text{if team } i \text{ did not play in time } t,
\end{aligned} \tag{4}$$

where $\theta_i = (C_i, B_i, A_i, M_{i,1}, \ldots, M_{i,n_M})'$ are the parameters for team i. The parameter $M_{i,m}$ denotes the relative strength of team i on map m with the value of $M_{i,1}$ set to 0 due to identifiability.

Alternative model suggests that the map choice does not significantly alter the result and thus the skill s is enough to denote the strength of a team. That would lead to team i defeating team j with probability

$$p_{i,j,t} = \frac{1}{1 + e^{-(S_{i,t} - S_{j,t})}} \tag{5}$$

with skill $S_{i,t}$ following recursion

$$\begin{aligned}
S_{i,t+1} &= C_i + B_i S_{i,t} + A_i \frac{e^{S_{k,t}}}{e^{S_{i,t}} + e^{S_{k,t}}} && \text{if team } i \text{ defeated team } k \text{ in time } t, \\
S_{i,t+1} &= C_i + B_i S_{i,t} - A_i \frac{e^{S_{i,t}}}{e^{S_{i,t}} + e^{S_{k,t}}} && \text{if team } i \text{ lost to team } k \text{ in time } t, \\
S_{i,t+1} &= S_{i,t} && \text{if team } i \text{ did not play in time } t.
\end{aligned} \tag{6}$$

The change in skill S_i of team i based on the skill S_j of the opponent can be observed in figs. 1 and 2.

The very last model we tried out was a static variant, where skill was fixed in time for each team, and the probability of team i winning over team j was again denoted by equation 6 with skill S_i of team i invariable in time, entering the model only as a constant.

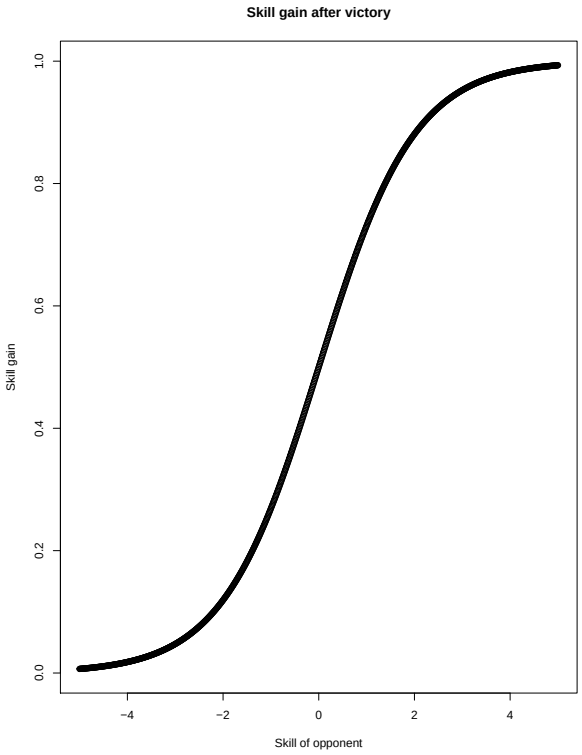


Figure 1: Skill increase after winning

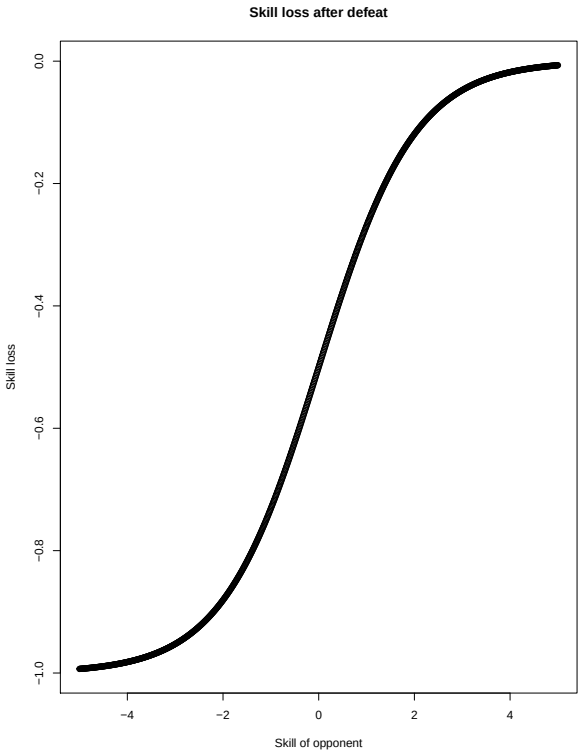


Figure 2: Skill decrease after losing

4 E-Sports Data

We analysed a little over 2000 matches from the 2015/2016 season that. Unlike soccer, which we took as inspiration for our analysis, there are no year-long leagues, but instead a larger amount of shorter tournaments which last about one week each. Also, because all of the matches are usually broadcast online in real time, very rarely are there more matches happening at the same time. Thanks to that, we can assign time t to each match which happens between exactly one match happening at $t-1$ and exactly one match at $t+1$.

As there are no fixed leagues and the standings and team rosters are more volatile compared to most traditional sports, we decided to select a sample of 16 most successful teams that play each other often, and used only the matches where both participants belong to our selected subset.

What makes eSport data different from sports data is the presence of maps. The matches are not played on an uniform field every time, but instead there is an predefined set of maps from which the teams select one to be played through a drafting process. The commonly employed strategy is to try and force a selection of a map the team performs well on. However, we presume that the optimal strategy would be to opt for a map where the disparity between effective strengths of the two participating teams is the largest.

For the purpose of our analysis, we only consider two possible outcomes for each match, either the victory of team i or victory of team j, representing the actual match, which goes on until a team reaches 15 points. In case the difference between scores is less than two by that time, the game enters overtime which lasts until one team pulls ahead.

5 Empirical Results

We ran all three of the proposed models and compared the results. The simplest model with no time variability was able to predict only 53.45% of the match results we tested the models on. This can be attributed to huge volatility of the eSports scene, where player transfers happen about ten times as often as they do in soccer and thus a model with static powers and no ability to adapt has problems and performed only slightly better than a coin flip.

The dynamic model with $A = 0.9, B = 0.1$ which considered variable skill S in time performed a bit better, successfully predicting 57.11% of match results from the validation set. Figures 3 and 4 show how skill evolves over time. In fig. 3 can be observed the difference between a team that was at the top for the whole season and a team signed a completely new roster during the season, which led to a dramatic increase in their winrate. Figure 4

shows that the skill for a average performing team, while dynamic is oscillating around a constant value.

Figure 3: Skill change over time - two top teams

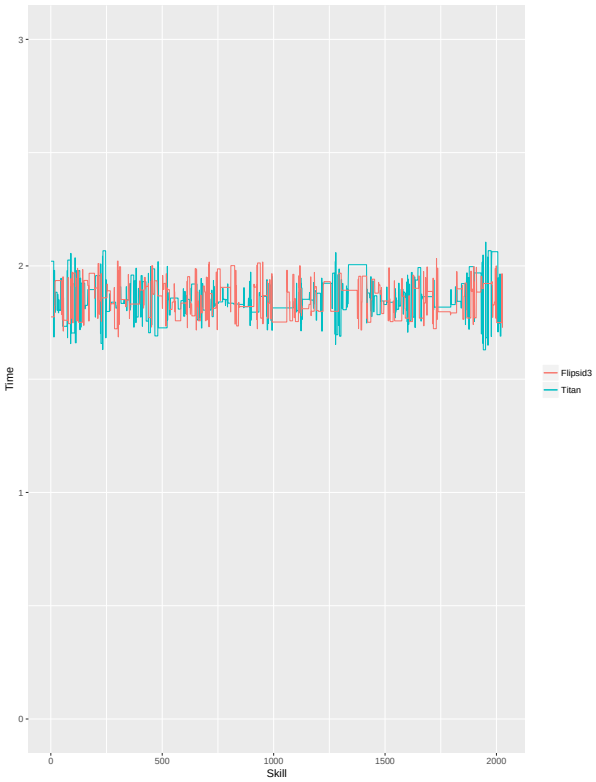


Figure 4: Skill change over time - two average teams

The second dynamic model, again with $A = 0.9, B = 0.1$, but including the relative map strength $M_{i,m}$ did not show significant improvement over the previous model, being able to predict correctly in 57.75% of the cases.

6 Conclusion

While the GAS models show potential in this area, the methodology needs to adapt in order to produce more precise forecasts. The environment is many times more volatile compared to traditional sports and adjustments need to be made to account for the frequent player transfers and changing form of teams, which both make prediction more complicated.

Several other factors impacting the result might prove interesting to include in future models, possibly different team strengths when on offence or defense as many recent soccer statisticians already use but the most important difference is still how dynamic the Counter-Strike environment is with about five times as many matches played per season when compared to soccer and very frequent transfers through the season as the game does not have a fixed transfer period. This implies that having variables dynamic in time is a must in order to maximize successful forecasting.

Acknowledgements

This paper was supported by the grant No. F4/21/2018 of the Internal Grant Agency of University of Economics, Prague.

References

[1] Avdulaj, K., and Barunik, J.: Are benefits from oil stocks diversification gone? new evidence from a dynamic copula and high frequency data. *Energy Economics* **51** (2015), 31 – 44.

[2] Cox, D. R.: Statistical Analysis of Time Series: Some Recent Developments. *Scandinavian Journal of Statistics* **8** (1981), 93–108.

[3] Creal, D., Koopman, S. J., and Lucas, A.: A General Framework for Observation Driven Time-Varying Parameter Models, 2008. URL `http://www.tinbergen.nl/discussionpaper/?paper=1416`.

[4] Creal, D., Koopman, S. J., and Lucas, A.: Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics* **28** (2013), 777–795.

[5] Dixon, M. J., and Coles, S. G.: Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **46** (1997), 265–280.

[6] Fahrmeir, L., and Tutz, G.: Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association* **89** (1994), 1438–1449.

[7] Harvey, A. C.: *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press, New York, 2013.

[8] Jenny, S. E., Manning, R. D., Keiper, M. C., and Olrich, T. W.: Virtual (ly) athletes: Where esports fit within the definition of sport. *Quest* **69** (2017), 1–18.

[9] Koopman, S. J., and Lit, R.: A dynamic bivariate poisson model for analysing and forecasting match results in the english premier league. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **178** (2015), 167–186.

[10] Koopman, S. J., Lucas, A., and Scharth, M.: Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models. *Review of Economics and Statistics* **98** (2016), 97–110.

[11] Koopman, S. J. S., Lit, R., et al.: *Forecasting Football Match Results in National League Competitions Using Score-Driven Time Series Models*. Technical report, Tinbergen Institute, 2017.

[12] Lucas, A., Schwaab, B., and Zhang, X.: Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics* **32** (2014), 271–284.

[13] Maher, M. J.: Modelling association football scores. *Statistica Neerlandica* **36** (1982), 109–118.

[14] Parshakov, P., and Zavertiaeva, M. A.: Success in esports: Does country matter? (2015).

[15] Reep, C., and Benjamin, B.: Skill and chance in association football. *Journal of the Royal Statistical Society. Series A (General)* **131** (1968), 581–585.

[16] Rue, H., and Salvesen, O.: Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)* **49** (2000), 399–418.

[17] Zhao, L., and van Wijnbergen, S.: Decision-making in incomplete markets with ambiguity a case study of a gas field acquisition. *Quantitative Finance* **17** (2017), 1759–1782.

Interval subeigenvectors of fuzzy matrices

Ján Plavka [1]

Abstract. A fuzzy algebra is a triple $(\mathbb{B}, \oplus, \otimes)$, where $(\mathbb{B}, \leq)$ is a nonempty, bounded, linearly ordered set and $a \oplus b = \max\{a, b\}$, $a \otimes b = \min\{a, b\}$ for $a, b \in \mathbb{B}$. A vector x is said to be a λ-eigenvector of a square matrix A if $A \otimes x = \lambda \otimes x$ for some $\lambda \in \mathbb{B}$. To solve subeigenproblem for some $\lambda \in \mathbb{B}$ means to find a solution x, called λ-subeigenvector, of $A \otimes x \leq \lambda \otimes x$. The values of vector or matrix inputs in practice are usually not exact numbers and in some intervals they can instead be considered as values. This paper investigates the properties of matrices and vectors with interval coefficients. In addition, a complete solution of the strong and the strong universal interval λ-subeigenproblem in fuzzy algebra is presented.

Keywords: fuzzy algebra, interval vector, subeigenvector

JEL classification: C44
AMS classification: 15A80, 15A18, 08A72

1 Introduction

Fuzzy algebra, the arithmetical operations $a \oplus b := \max(a, b)$ and $a \otimes b := \min(a, b)$ are defined over a linearly ordered set, can be applied to various problems related to scheduling and optimization. Fuzzy algebra is a part of the so-called tropical mathematics, and it has a wide scope for application and it offers an interesting contribution to the mathematical theory.

As an economic motivation for the research of fuzzy algebra can be considered adapting max–plus interaction systems, [1], [2], i.e., on an assembly line a man cannot begin a new assembly until, say, two inter-locking sub-assemblies have arrived from different sources with independent production rates. A natural way of describing such system is to label the machines (entities), e.g., $1, 2, \ldots, n$, and to describe the interferences by recurrence relations

$$x_i(r+1) = \max(x_1(r) + a_{1i},\ x_2(r) + a_{2i}, \ldots,\ x_n(r) + a_{ni}),\ i \in \{1, 2, \ldots, n\}.$$

The last formula expresses the fact that machine i must wait to begin its $r+1$st cycle until machines $j = 1, \ldots, n$ have finished their rth cycle, the symbol $x_i(r)$ denotes the starting time of the rth cycle of machine i, and a_{ij} is corresponding activity duration.

Denote $a \oplus b = \max(a, b)$, and $a \otimes b = a + b$ for $a, b \in \mathbb{R}$ and extend this pair of operations to matrices and vectors in the same way as in conventional linear algebra.

By generalization of the discrete-event dynamic system (eigenproblem) mentioned above and using the above notation we obtain a description of the steady state system (for some constant λ, the interval between the beginnings of consecutive cycles on every machine has to be λ, i.e., $x(r+1) = A \otimes x(r)$ and $x(r+1) = \lambda \otimes x(r)$) in vector-matrix notation as follows:

$$A \otimes x = \lambda \otimes x.$$

For simplicity A does not change from stage to stage. 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 system. If instead of the equality in $A \otimes x = \lambda \otimes x$ we ask for inequality $\leq$, then we speak about a subeigenproblem.

For a given matrix $A \in \mathbb{B}(n, n)$ in fuzzy algebra, the eigenproblem consists of finding a value $\lambda \in \mathbb{B}$ and a vector $x \in \mathbb{B}(n)$ such that the equation $A \otimes x = \lambda \otimes x$ holds true. The eigenproblem in fuzzy algebra has been studied by many authors. Interesting results were found in describing the structure of the eigenspace (the set of all eigenvectors) [17] and several algorithms for computing the largest eigenvector of a given matrix have been suggested, see for example [16].

A vector x is said to be a λ-eigenvector of a square matrix A if $A \otimes x = \lambda \otimes x$ for some $\lambda \in \mathbb{B}$. To solve subeigenproblem for some $\lambda \in \mathbb{B}$ means to find a solution x of $A \otimes x \leq \lambda \otimes x$, x is called λ-subeigenvector. In [15] the properties of λ-subeigenvectors are described, the values λ associated with λ-subeigenvectors are characterized and efficient algorithms for checking all equivalent conditions are introduced.

In the present paper, we consider an interval version of this condition and investigate the properties of matrices and vectors with interval coefficients, i.e., we describe λ-subeigenvectors for matrices $A \in \mathbb{B}(n, n)$ and an interval

[1] Technical University in Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, Jan.Plavka@tuke.sk

$\boldsymbol{X} = [\underline{x}, \overline{x}] := \{x \in \mathbb{B}(n);\ \underline{x} \leq x \leq \overline{x}\}$. In addition, a complete solution of the strong and the strong universal interval subeigenproblem in fuzzy algebra is presented.

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 fuzzy algebra, leading to the discussion of structures of subeigenproblem in Section 3. Section 4 is devoted to the main results of the paper which characterizes interval matrices with the strong and the strong universal interval λ-subeigenvectors.

2 Background of the problem

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. The set of naturals (naturals with zero) is denoted by $\mathbb{N}$ ($\mathbb{N}_0$). For given naturals $n, m \in \mathbb{N}$, we use notations N and M for the set of all smaller or equal natural numbers, i.e., $N = \{1, 2, \ldots, n\}$ and $M = \{1, 2, \ldots, m\}$, respectively. The set of $n \times m$ matrices over $\mathbb{B}$ is denoted by $\mathbb{B}(n, m)$, specially the set of $n \times 1$ vectors over $\mathbb{B}$ is denoted by $\mathbb{B}(n)$ and for $\alpha \in \mathbb{B}$ a constant vector is denoted by $\alpha^\star = (\alpha, \ldots, \alpha)^T$.

A fuzzy algebra is a triple $(\mathbb{B}, \oplus, \otimes)$, where $a \oplus b = \max(a, b)$ and $a \otimes b = \min(a, b)$.

The operations $\oplus, \otimes$ are extended to the matrix-vector algebra over $\mathbb{B}$ by the direct analogy to the conventional linear algebra.

For $A \in \mathbb{B}(n, n)$, $C \in \mathbb{B}(n, n)$ we write $A \leq C$ ($A < C$) if $a_{ij} \leq c_{ij}$ ($a_{ij} < c_{ij}$) holds true for all $i, j \in N$. Similarly, for $x = (x_1, \ldots, x_n)^T \in \mathbb{B}(n)$ and $y = (y_1, \ldots, y_n)^T \in \mathbb{B}(n)$ we write $x \leq y$ ($x < y$) if $x_i \leq y_i$ ($x_i < y_i$) for each $i \in N$.

The rth power of a matrix A is denoted by A^r with elements a_{ij}^r.

By a digraph we understand a pair $G = (V_G, E_G)$, where V_G is a non-empty finite set, called the node set, and $E_G \subseteq V_G \times V_G$, called the arc set. A digraph G' is a subdigraph of G, if $V_{G'} \subseteq V_G$ and $E_{G'} \subseteq E_G$. A path in G is the sequence of nodes $p = (v_0, v_1, \ldots, v_l)$ such that $(v_{k-1}, v_k) \in E_G$ for all $k = 1, 2, \ldots, l$ (denoted as (v_0, v_l)–path). The number $l(p) \geq 0$ is called the length of p. If $v_0 = v_l$, then p is called a cycle. A cycle is elementary if all nodes except the terminal node are distinct. A digraph is called strongly connected if any two distinct nodes of G are contained in a common cycle.

By a *strongly connected component* $\mathcal{K}$ of $G = (N_G, E_G)$ we mean a subdigraph $\mathcal{K}$ generated by a non-empty subset $K \subseteq N$ such that any two distinct nodes $i, j \in K$ are contained in a common cycle and 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 $\mathcal{K}$.

The set of all non-trivial strongly connected components of G is denoted by SCC(G).

The $eigenproblem$ in fuzzy algebra is formulated as follows: Given $A \in \mathbb{B}(n, n)$ and the number $\lambda \in \mathbb{B}$, find $x \in \mathbb{B}(n)$ satisfying

$$A \otimes x = \lambda \otimes x,$$

n-tuple $x \in \mathbb{B}(n)$ is the called *eigenvector* of A and $\lambda \in \mathbb{B}$ is *eigenvalue* of A.

The *eigenspace* $V(A, \lambda)$ is defined as the set of all eigenvectors of A with associated eigenvalue λ, i.e., $V(A, \lambda) = \{x \in \mathbb{B}(n);\ A \otimes x = \lambda \otimes x\}$.

In case $\lambda = I$ let us denote $V(A, I)$ by the abbreviation $V(A)$.

Notice that if $x \in V(A)$, i.e., $A \otimes x = x$, then for each $\lambda \in \mathbb{B}$ the equalities

$$A \otimes (\lambda \otimes x) = \lambda \otimes (A \otimes x) = \lambda \otimes x = \lambda \otimes (\lambda \otimes x)$$

hold true. Thus, every $\lambda \in \mathbb{B}$ is an eigenvalue of A.

Let $A \in \mathbb{B}(n, n)$ be given. The symbol A^+ stands for $\lim_{j \to \infty} (A \oplus A^2 \oplus \cdots \oplus A^j)$ and $A^* = E \oplus A^+$, where E is unit matrix. $A^* = (a_{ij}^*)$ is called the strong transitive closure of A.

There is a well-known connection between the entries in powers of matrices and paths in associated digraphs: the (i, j)th entry a_{ij}^k in A^k is equal to the maximum of weights of paths from $\mathcal{P}_{ij}^k$, where $\mathcal{P}_{ij}^k$ is the set of all paths of length k beginning at node i and ending at node j. If $\mathcal{P}_{ij}$ denotes the set of all paths from i to j, then $a_{ij}^* = \max\{a_{ij}^k;\ k = 1, 2, \ldots\}$ is the maximum weight of a path from $\mathcal{P}_{ij}$ and a_{jj}^* is the maximum weight of all

cycles containing node j.

For every matrix $A \in \mathbb{B}(n,n)$ denote

$$c_i(A) = \bigoplus_{j\in N} a_{ij},\ c(A) = \bigotimes_{i\in N} c_i(A),\ c_A = (c_1(A), c_2(A), \ldots, c_n(A))^T \in \mathbb{B}(n),$$

$$\underline{\lambda}(A)(=\underline{\lambda}) = \bigotimes_{i\in N} a_{ii} \text{ and } \overline{\lambda}(A)(=\overline{\lambda}) = \bigoplus_{i,j\in N} a_{ij}\ .$$

Notice that both operations in fuzzy algebra are idempotent, hence the range of mappings $\max$ and $\min$ is $\mathbb{B}$ and no new numbers are created in the process of generating matrix-vector products.

A matrix (vector) is called *binary* if $a_{ij} \in \{O, I\}$ ($x_j \in \{O, I\}$) for each $i, j \in N$.

Definition 1. Let $A \in \mathbb{B}(n,n)$ be a binary matrix and $x \in \mathbb{B}(n)$ be a binary vector. Then by $G(A) = (V_{G(A)}, E_{G(A)})$ we understand the digraph with $V_{G(A)} = N$, $E_{G(A)} = \{(i,j);\ a_{ij} = I\}$.

Definition 2. For $A \in \mathbb{B}(n,n)$ and $h \in \mathbb{B}$, the *threshold matrix* $A_{(h)}$ ($A_{(h^+)}$) corresponding to the *threshold* h is a binary matrix of the same type as A defined as follows:

$$(a_{(h)})_{ij} = \begin{cases} I, & \text{if } a_{ij} \geq h, \\ O, & \text{otherwise,} \end{cases} \qquad \left((a_{(h^+)})_{ij} = \begin{cases} I, & \text{if } a_{ij} > h, \\ O, & \text{otherwise} \end{cases} \right). \tag{1}$$

The associated digraph $G(A_{(h)})$ ($G(A_{(h^+)})$) will be called the *threshold digraph* corresponding to the threshold h (h^+) .

3 Structures of subeigenvectors and supereigenvectors

For $A \in \mathbb{B}(n,n)$ and $\lambda \in \mathbb{B}$ we denote

$$V_*(A,\lambda) = \{x \in \mathbb{B}(n);\ A \otimes x \leq \lambda \otimes x\},$$

$$V^*(A,\lambda) = \{x \in \mathbb{B}(n);\ A \otimes x \geq \lambda \otimes x\}$$

or just $V_*(A)$, $V^*(A)$ for $\lambda = I$.
Notice that the sets $V_*(A,\lambda)$, $V^*(A,\lambda)$ are not empty for each A and for each λ because there is a greatest eigenvector belonging to $V_*(A,\lambda) \cap V^*(A,\lambda)$.

If $A \in \mathbb{B}(n,n)$ and $\lambda \in \mathbb{B}$ then a vector $x \in \mathbb{B}(n)$ satisfying $A \otimes x \leq \lambda \otimes x$ ($A \otimes x \geq \lambda \otimes x$) is called a λ-subeigenvector (λ-supereigenvector) of A with associated subeigenvalue (supereigenvalue) λ and $V_*(A,\lambda)$ ($V^*(A,\lambda)$) is called λ-subeigenspace (λ-supereigenspace).

From the definitions of $V_*(A,\lambda)$ and $V^*(A,\lambda)$ the next simple implications follow.

$$x \in V_*(A,\lambda) \Rightarrow (\forall k \in \mathbb{N}_0)[A^k \otimes x \geq A^{k+1} \otimes x], \tag{2}$$

$$x \in V^*(A,\lambda) \Rightarrow (\forall k \in \mathbb{N}_0)[A^k \otimes x \leq A^{k+1} \otimes x]. \tag{3}$$

Lemma 1. [15] If $x \in V_*(A) \cup V^*(A)$, then $A^n \otimes x \in V(A)$.

Theorem 2. [15] *If $A \in \mathbb{B}(n,n)$ is a given matrix, $\lambda \in \mathbb{B}$ is a given value,* $\mathrm{SCC}(G(A_{(\lambda)})) = \bigcup_{i=1}^{p} K_i^{\lambda}(V_i^{\lambda}, E_i^{\lambda})$ *and* $\mathrm{SCC}(G(A_{(\lambda^+)})) = \bigcup_{i=1}^{m} K_i^{\lambda^+}(V_i^{\lambda^+}, E_i^{\lambda^+})$*, then*

(i) $V_*(A,\lambda) = V(A,\lambda) \Rightarrow \lambda \leq \underline{\lambda} \vee \bigcup_{i=1}^{m} V_i^{\lambda} = N.$

(ii) $\lambda \leq \underline{\lambda} \vee \bigcup_{i=1}^{p} V_i^{\lambda^+} = N \Rightarrow V_*(A,\lambda) = V(A,\lambda).$

(iii) If $\lambda \in \mathbb{B} \setminus \{a_{ij};\ i,j \in N\}$*, then* $V_*(A,\lambda) = V(A,\lambda) \Leftrightarrow \lambda \leq \underline{\lambda} \vee \bigcup_{i=1}^{m} V_i^{\lambda^+} = N.$

For $A \in \mathbb{B}(n,n)$ denote $\bigotimes_{(A\otimes c_A)_i<(c_A)_i} (A \otimes c_A)_i$ by $\tilde{\lambda}(A)$ or just $\tilde{\lambda}$.

Theorem 3. [15] *If $c_A \leq x \leq \overline{\lambda}^{\star}$, then the following hold true:*

(i) If $\lambda \leq \tilde{\lambda}$, then $x \in V^*(A,\lambda) \Leftrightarrow \lambda \otimes x = \lambda \otimes c_A$.

(ii) If $\tilde{\lambda} < \lambda$, then $V^*(A,\lambda) = \emptyset$.

4 Interval versions of λ-subeigenvectors

Similarly to [5], [9], [10], [11], [14], we define interval matrix with bounds $\underline{A}, \overline{A} \in \mathbb{B}(n,n)$ and interval vector with bounds $\underline{x}, \overline{x} \in \mathbb{B}(n)$ as follows

$$\boldsymbol{A} = [\underline{A}, \overline{A}] = \left\{ A \in \mathbb{B}(n,n);\ \underline{A} \leq A \leq \overline{A} \right\},$$

$$\boldsymbol{X} = [\underline{x}, \overline{x}] = \left\{ x \in \mathbb{B}(n);\ \underline{x} \leq x \leq \overline{x} \right\}.$$

We assume in this section that an interval matrix $\boldsymbol{A} = [\underline{A}, \overline{A}]$, $\lambda \in \mathbb{B}$ and an interval vector $\boldsymbol{X} = [\underline{x}, \overline{x}]$ are fixed. The interval subeigenproblem for $\boldsymbol{A}$, $\lambda \in \mathbb{B}$ and $\boldsymbol{X}$ consists in recognizing whether $A \otimes x \leq \lambda \otimes x$ holds true for $A \in \boldsymbol{A}$, $x \in \boldsymbol{X}$. In dependence on the applied quantifiers, we get two types of interval λ-subeigenvectors.

Definition 3. If interval matrix $\boldsymbol{A}$ and $\lambda \in \mathbb{B}$ are given, then interval vector $\boldsymbol{X}$ is called
- *strong λ-subeigenvector* of $\boldsymbol{A}$ if $(\forall A \in \boldsymbol{A})(\forall x \in \boldsymbol{X})[\, A \otimes x \leq \lambda \otimes x \,]$,
- *strong universal λ-subeigenvector* of $\boldsymbol{A}$ if $(\exists x \in \boldsymbol{X})(\forall A \in \boldsymbol{A})\, [A \otimes x \leq \lambda \otimes x\,]$.

The remainder of this paper assumes we are given an interval matrix $\boldsymbol{A} = [\underline{A}, \overline{A}]$ and an interval vector $\boldsymbol{X} = [\underline{x}, \overline{x}]$. For each pair of indices $i, j \in N$, we define $\tilde{A}^{(ij)} \in \mathbb{B}(n,n)$ and $\tilde{x}^{(i)} \in \mathbb{B}(n)$ by putting for every $k, l \in N$,

$$\tilde{a}_{kl}^{(ij)} = \begin{cases} \overline{a}_{ij}, & \text{for } k = i,\ l = j \\ \underline{a}_{kl}, & \text{otherwise} \end{cases}, \qquad \tilde{x}_k^{(i)} = \begin{cases} \overline{x}_i, & \text{for } k = i \\ \underline{x}_k, & \text{otherwise} \end{cases}.$$

The next lemma says that every matrix in $\boldsymbol{A}$ can be written as a max-min linear combination of *generating* matrices ("generators," for short) $\tilde{A}^{(ij)}$ with $i,\ j \in N$. Similarly, every vector in $\boldsymbol{X}$ is equal to a max-min linear combination of generators $\tilde{x}^{(i)}$ with $i \in N$.

Lemma 4. Let $x \in \mathbf{X}$ and $A \in \mathbf{A}$. Then
(i) $x = \bigoplus_{i=1}^{n} \beta_i \otimes \tilde{x}^{(i)}$ for some numbers $\beta_i \in \mathbb{B}$,
(ii) $A = \bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes \tilde{A}^{(ij)}$ for some numbers $\alpha_{ij} \in \mathbb{B}$.

Proof. Let us suppose that $x \in \mathbf{X}$ and $A \in \mathbf{A}$. It is easily seen that the assertions follow for $\beta_i = x_i$ and $\alpha_{ij} = a_{ij}$. □

4.1 Strong interval λ-subeigenvectors

Theorem 5. *Let $\boldsymbol{A}$, $\boldsymbol{X}$ and $\lambda \in \mathbb{B}$ be given. The interval vector $\boldsymbol{X}$ is a strong λ-subeigenvector of $\boldsymbol{A}$ if and only if $(\forall A \in \boldsymbol{A})(\forall k \in N)[A \otimes \tilde{x}^{(k)} \leq \lambda \otimes \tilde{x}^{(k)}]$.*

Proof. Assume that there is $\lambda \in \mathbb{B}$ such that $(\forall A \in \boldsymbol{A})(\forall k \in N)[A \otimes \tilde{x}^{(k)} \leq \lambda \otimes \tilde{x}^{(k)}]$. Then by Lemma 4(i) for arbitrary $x \in \boldsymbol{X}$ we get

$$A \otimes x = A \otimes \bigoplus_{k=1}^{n} \beta_k \otimes \tilde{x}^{(k)} = \bigoplus_{k=1}^{n} \beta_k \otimes (A \otimes \tilde{x}^{(k)}) \leq \bigoplus_{k=1}^{n} \beta_k \otimes (\lambda \otimes \tilde{x}^{(k)}) = \lambda \otimes x.$$

The converse implication trivially follows. □

Theorem 6. *Let $\boldsymbol{A}$, $\boldsymbol{X}$ and $\lambda \in \mathbb{B}$ be given. The interval vector $\boldsymbol{X}$ is a strong λ-subeigenvector of $\boldsymbol{A}$ if and only if $(\forall i, j, k \in N)[\tilde{A}^{(ij)} \otimes \tilde{x}^{(k)} \leq \lambda \otimes \tilde{x}^{(k)}]$.*

Proof. Let us assume that $\tilde{A}^{(ij)} \otimes \tilde{x}^{(k)} \leq \lambda \otimes \tilde{x}^{(k)}$ for all $i,j,k \in N$. Then by Lemma 4 for arbitrary $x \in \boldsymbol{X}$ and $A \in \mathbb{B}(n,n)$ we get

$$A \otimes x = (\bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes \tilde{A}^{(ij)}) \otimes (\bigoplus_{k=1}^{n} \beta_k \otimes \tilde{x}^{(k)}) = \bigoplus_{i,j=1}^{n} \bigoplus_{k=1}^{n} (\alpha_{ij} \otimes \tilde{A}^{(ij)}) \otimes (\beta_k \otimes \tilde{x}^{(k)}) =$$

$$\bigoplus_{i,j=1}^{n} \alpha_{ij} \otimes \bigoplus_{k=1}^{n} (\beta_k \otimes \tilde{A}^{(ij)}) \otimes \tilde{x}^{(k)} = \bigoplus_{k=1}^{n} \beta_k \otimes (\tilde{A}^{(ij)} \otimes \tilde{x}^{(k)}) \leq$$

$$\bigoplus_{k=1}^{n} \beta_k \otimes \lambda \otimes \tilde{x}^{(k)} = \lambda \otimes \bigoplus_{k=1}^{n} \beta_k \otimes \tilde{x}^{(k)} = \lambda \otimes x.$$

The converse implication is trivial.

□

Theorem 7. *Suppose given a matrix $\boldsymbol{A}$, λ and an interval vector $\boldsymbol{X} = [\underline{x}, \overline{x}]$. The recognition problem of whether $\boldsymbol{X}$ is strongly λ-subeigenproblem is solvable in $O(n^6)$ time.*

Proof. According to Theorem 6, the recognition problem whether $\boldsymbol{X}$ is strong interval λ-subeigenproblem is equivalent to recognizing whether $(\forall i,j,k \in N)[\tilde{A}^{(ij)} \otimes \tilde{x}^{(k)} \leq \lambda \otimes \tilde{x}^{(k)}]$ hold true. The computation of a system $C \otimes y \leq D \otimes y$ needs $O(mn \cdot \min(m,n))$ time (see [4]), where $C, D \in \mathbb{B}(m,n)$. Therefore, the computation of $\tilde{A}^{(ij)} \otimes \tilde{x}^{(k)} \leq \lambda \otimes \tilde{x}^{(k)}]$ is done in $O(n^3 \cdot n^3) = O(n^6)$ time. □

4.2 Strong universal interval λ-subeigenvectors

Denote the block matrices $C \in \mathbb{B}(n^2, n)$, $D \in \mathbb{B}(n^2, n)$ as follows

$$C = \begin{pmatrix} \tilde{A}^{(11)} \otimes \tilde{x}^{(1)} & \dots & \tilde{A}^{(11)} \otimes \tilde{x}^{(n)} \\ \tilde{A}^{(12)} \otimes \tilde{x}^{(1)} & \dots & \tilde{A}^{(12)} \otimes \tilde{x}^{(n)} \\ \vdots & \vdots & \vdots \\ \tilde{A}^{(nn)} \otimes \tilde{x}^{(1)} & \dots & \tilde{A}^{(nn)} \otimes \tilde{x}^{(n)} \end{pmatrix}, \quad D = \begin{pmatrix} \tilde{x}^{(1)} & \dots & \tilde{x}^{(n)} \\ \tilde{x}^{(1)} & \dots & \tilde{x}^{(n)} \\ \vdots & \vdots & \vdots \\ \tilde{x}^{(1)} & \dots & \tilde{x}^{(n)} \end{pmatrix}.$$

Theorem 8. *Let $\boldsymbol{A}$, $\boldsymbol{X}$ and $\lambda \in \mathbb{B}$ be given. Then $\boldsymbol{X}$ is a strong universal λ-subeigenvector of $\boldsymbol{A}$ if and only if the two-side system $C \otimes \beta \leq \lambda \otimes D \otimes \beta$ has a solution $\beta = (\beta_1, \dots, \beta_n)^T$ such that $\underline{x} \leq \bigoplus_{i=1}^{n} \beta_i \otimes \tilde{x}^i \leq \overline{x}$.*

Proof. By Lemma 4(i), if $\underline{x}_j \leq \beta_j \leq \overline{x}_j$ for $j \in N$ then $x = \bigoplus_{j=1}^{k} \beta_j \otimes x^{(j)}$ belongs to $\boldsymbol{X}$, and if $x \in \boldsymbol{X}$ then we can find $\underline{x}_j \leq \beta_j \leq \overline{x}_j$ for $j \in N$ such that $x = \bigoplus_{j=1}^{k} \beta_j \otimes x^{(j)}$.

We also have that the system of inequalities $C \otimes \beta \leq \lambda \otimes D \otimes \beta$ has a solution $\beta = (\beta_1, \dots, \beta_n)^T$ with $\underline{x} \leq \bigoplus_{i=1}^{n} \beta_i \otimes \tilde{x}^i \leq \overline{x}$ if and only if the following equivalences for an arbitrary $i,j \in N$ hold true:

$$C \otimes \beta \leq \lambda \otimes D \otimes \beta \Leftrightarrow$$

$$(\forall i,j \in N)[\tilde{A}^{(ij)} \otimes \bigoplus_{k \in N} x^{(k)} \otimes \beta_k \leq \lambda \otimes \bigoplus_{k \in N} x^{(k)} \otimes \beta_k] \Leftrightarrow$$

$$(\exists x \in \boldsymbol{X})(\forall i,j \in N)[\tilde{A}^{(ij)} \otimes x \leq \lambda \otimes x]$$

and hence by Lemma 4(ii) we obtain

$$(\exists x \in \boldsymbol{X})[\bigoplus_{i,j \in N} \tilde{A}^{(ij)} \otimes x \leq \lambda \otimes x \Leftrightarrow (\exists x \in \boldsymbol{X})(\forall A \in \boldsymbol{A})[A \otimes x \leq \lambda \otimes x].$$

□

Theorem 9. *Suppose given a matrix $\boldsymbol{A}$, λ and an interval vector $\boldsymbol{X} = [\underline{x}, \overline{x}]$. The recognition problem of whether $\boldsymbol{X}$ is strong universal λ-subeigenproblem is solvable in $O(n^5)$ time.*

Proof. According to Theorem 8, the recognition problem whether $\boldsymbol{X}$ is strong universal λ-subeigenproblem is equivalent to recognizing whether system $C \otimes \beta \leq \lambda \otimes D \otimes \beta$ has a solution $\beta = (\beta_1, \ldots, \beta_n)^T$ such that $\underline{x} \leq \bigoplus_{i=1}^{n} \beta_i \otimes \tilde{x}^i \leq \overline{x}$. The computation of a system $C \otimes y \leq D \otimes y$ needs $O(mn \cdot \min(m,n))$ time (see [4]), where $C, D \in \mathbb{B}(m,n)$. Therefore, the computation of $C \otimes \beta \leq \lambda \otimes D \otimes \beta$ is done in $O((n^2)^2 \cdot n) = O(n^5)$ time. □

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] Gavalec, M., Gad, M. and Zimmermann, K.: Optimization Problems under (max, min)- Linear Equations and / or Inequality Constraints, Journal of Mathematical Sciences 193, 5 (2013), 645–658.

[5] Molnárová, M., Myšková, H. and Plavka, J.: The robustness of interval fuzzy matrices, *Linear Algebra and Its Applications* 438(8) (2013) 3350–3364.

[6] Molnárová, M.: Generalized matrix period in max-plus algebra, *Linear Algebra and Its Applications* 404 (2005) 345–366.

[7] Molnárová, M.: Computational complexity of Nachtigall's representation, *Optimization* 52(1) (2003) 93–104.

[8] Molnárová, M. and Pribiš, J.: Matrix period in max-algebra, *Discrete Applied Mathematics* 103(1-3) (2000) 167–175.

[9] Myšková, H.: Robustness of Interval Toeplitz Matrices in Fuzzy Algebra, Acta Electrotechnica et Informatica, Vol. 12, No. 4, 2012, 56–60.

[10] Myšková, H.: Interval eigenvectors of circulant matrices in fuzzy algebra *Acta Electrotechnica et Informatica*. Vol. 12, No. 3, 2012, 57–61.

[11] Myšková, H.: Weak stability of interval orbits of circulant matrices in fuzzy algebra. *Acta Electrotechnica et Informatica*, Vol. 12, No. 3, 2012, 51–56.

[12] Plavka J., and Szabó, P.: On the λ-robustness of matrices over fuzzy algebra, *Discrete Applied Mathematics* 159(5) (2011), 381–388.

[13] Plavka, J.: On the weak robustness of fuzzy matrices, *Kybernetika* 49(1) (2013), 128–140.

[14] Plavka, J., and Sergeev, S.: Characterizing matrices with X-simple image eigenspace in max-min semiring, *Kybernetika* 52(4) (2016), 497–513.

[15] Plavka, J.: Subeigenvectors and supereigenvectors of fuzzy matrices, Fuzzy Sets and Systems 339 (2018) 62–72.

[16] Tan, Yi-Jia: On the eigenproblem of matrices over distributive lattices, *Linear Algebra and its Applications* 374 (2003), 87–106.

[17] Tomášková, H: Eigenproblem for circulant matrices in max-plus algebra. In: Proceedings of the 29th Conference on Mathematical Methods, Computational Techniques, Intelligent Systems (MAMECTIS-29), 2010.

[18] Zimmernann, K.: Extremální algebra (in Czech), *Ekon. ústav ČSAV* Praha, 1976.

Multi-criteria portfolio selection with fuzzy weights of the criteria – comparison of selected approaches

Ewa Pośpiech[1]

Abstract. The issue of portfolio selection can be viewed as a multi-criteria problem. This approach is justified due to the variety of characteristics on the basis of which a company shall be assessed. The choice of characteristics which are the evaluation criteria, may significantly determine the results. Besides, the application of multi-criteria methods in portfolio selection is connected with other subjective decisions and ambiguous assessments which also affect the results. One of such subjective decision is the choice of the values of the criteria weights. The weights can be specified as one value but they can also be given as ranges of values that indicate the smallest and the largest possible value of the weight of the criterion. The aim of the article is to compare results of the ordering and the selection of shares that support building a portfolio in the situation of ambiguous weights. Therefore, in the paper, fuzzy numbers concerning the criteria weights are applied and selected multi-criteria methods are used.

Keywords: fuzzy weights of criteria, multi-criteria methods, portfolio selection.

JEL Classification: C44, C61, G11
AMS Classification: 62C86, 90B50

1 Introduction

In each multi-criteria issue there is a problem of selection criteria weights. This is an important element of the considerations that determines the solution. When selecting the criteria weights the decision maker is guided by his own preferences and experiences, but should also take into account the suggestions of experts. This approach makes that the weights may not be specified explicitly. This means that in the considered issue the weights may be expressed not as a single value but can be describe as a range of values. Therefore, we have to deal with the ambiguity of weights which should be properly taken into account in the deliberations.

The shares selection to the portfolio is a complex issue. There are many ways to support the selection of shares – preferred are those shares which are characterized by a high rate of return and low risk. However, in practice, the rate of return and risk are in general positively correlated. Therefore, additional analyses with other tools are necessary (e.g. ratio analysis or fundamental analysis tools). Proponents of fundamental analysis examine the characteristics that describe the economic and financial condition of companies. This approach makes the issue of portfolio selection can be dealt as a multi-criteria problem in which the fundamental and market characteristics are the criteria of selection. In multi-criteria problems with ambiguous weights you can use fuzzy modelling tools. In the paper, triangular fuzzy numbers are used that reflect the ambiguous weights of criteria. The subject matter of the article is a continuation of earlier analyses aimed at applying multi-criteria methods to the selection of a portfolio taking into account the so-called soft modelling [7], [8], [9], [10].

The aim of the article is to compare the results of ordering and grouping of listed companies constituting the basis of portfolio selection (with the use of Markowitz model) obtained using selected multi-criteria methods with ambiguous weights. The research hypothesis assumes more profitable portfolios selected on the basis of fuzzy approach. The article consists of two parts – the first one presents a brief description of multi-criteria methods used in the research, while in the second are presented the results of the analyses and conclusions.

2 Selected multi-criteria methods

The analyzes used multi-criteria SAW, TOPSIS and PROMETHEE II methods. The choice of the SAW method was caused by the fact that this method was developed in a fuzzy version with triangular fuzzy numbers [14], while the choice of other methods was dictated by previously conducted studies [9].

[1] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, 1 Maja 50, 40-287 Katowice, ewa.pospiech@ue.katowice.pl

2.1 Fuzzy Simple Additive Weighting Method (FSAW)

The SAW Method (Simple Additive Weighting Method) is one of the most familiar and intuitive multi-criteria methods. It enables the modelling of preferences using the additive linear function. Let $a_i^{(k)}$ means the assessment of variant i, $i = 1, \ldots, m$ through the prism of the criterion k, $k = 1, \ldots, n$. The matrix $\mathbf{X} = [x_{ik}]$ of normalised values is built. The matrix elements are defined as follows [2], [14]:

$$x_{ik} = \frac{a_i^{(k)} - \min\limits_{1 \le l \le m} a_l^{(k)}}{\max\limits_{1 \le l \le m} a_l^{(k)} - \min\limits_{1 \le l \le m} a_l^{(k)}}, \text{ when the criterion is maximized} \tag{1}$$

$$x_{ik} = \frac{\max\limits_{1 \le l \le m} a_l^{(k)} - a_i^{(k)}}{\max\limits_{1 \le l \le m} a_l^{(k)} - \min\limits_{1 \le l \le m} a_l^{(k)}}, \text{ when the criterion is minimized.} \tag{2}$$

The assessment of decision variant i is carried out on the basis of the value of p_i coefficient; is expressed by the formula (the higher the value, the variant is higher ranked in the ranking):

$$p_i = \sum_{k=1}^{n} w_k x_{ik} \ . \tag{3}$$

This method was also developed in fuzzy version (FSAW) [14], [15]. In this approach weights are treated as the following triangular fuzzy numbers: (l_k, m_k, u_k), where: l_k – the smallest value of the weight, u_k – the largest value of the weight, m_k – the most appropriate value of the weight, moreover $\sum_{k=1}^{n} m_k = 1$. In this situation, the assessment will be based on the parameter $\widetilde{p}_i$ of the following form:

$$\widetilde{p}_i = \sum_{k=1}^{n} \widetilde{w}_k x_{ik} \ , \tag{4}$$

where $\widetilde{w}_k$ specifies the weight as a triangular fuzzy number. The variants are ordered according to descending values of the number p_i^D which is estimated using the Centre of Gravity method. The number is calculated from the formula:

$$p_i^D = \sum_{k=1}^{n} \frac{l_k + m_k + u_k}{3} x_{ik} \ . \tag{5}$$

2.2 TOPSIS with fuzzy weights

The TOPSIS method allows to put the variants in order by comparing them with reference points (ideal and anti-ideal one) – the closer to the ideal point and farther from the anti-ideal point, the higher position in the ranking the variant takes. A detailed description of the method can be found e.g. in [3], [4], [14]. The basis for discussion is the appropriate normalized matrix $\mathbf{X} = [\hat{x}_{ik}]$, for $i = 1, 2, \ldots, m$, $k = 1, 2, \ldots, n$; the elements of this matrix are then multiplied by the weights. The matrix is obtained as $\mathbf{Z} = [w_k \hat{x}_{ik}] = [v_{ik}]$, where w_k, $k = 1, 2, \ldots, n$, are the weights assigned to the criteria. For these elements the weighted points are designated: the ideal v_k^+ and the anti-ideal v_k^-. Then, using for example the Euclidean distance, the distance between the variants and the designated reference points is calculated. These distances are marked with the following symbols: d_i^+ and d_i^- – on the basis of these values the indicator S_i is built:

$$S_i = \frac{d_i^-}{d_i^+ + d_i^-}, \quad i = 1, 2, \ldots, m. \tag{6}$$

The indicator takes values from 0 to 1, the closer to 1 the value of S_i, the higher the position of the variant in ranking. In addition to the standard version of the TOPSIS method in which the weighted normalised decision matrix $\mathbf{Z} = [w_k \hat{x}_{ik}]$, $k = 1, 2, \ldots, n$, was used, the case in which the weights are expressed as presented above triangular fuzzy numbers (l_k, m_k, u_k). The matrix $\mathbf{Z}$ contains the following triangular fuzzy numbers:

$$\mathbf{Z} = [(l_k \hat{x}_{ik}, m_k \hat{x}_{ik}, u_k \hat{x}_{ik})], \ k = 1, 2, \ldots, n, \tag{7}$$

while further analysis uses a matrix $\hat{\mathbf{Z}} = [\hat{z}_{ik}]$, $k = 1, 2, \ldots, n$, which elements determined by means of the Centre of Gravity method, are as follows:

$$\hat{z}_{ik} = \frac{l_k \hat{x}_{ik} + m_k \hat{x}_{ik} + u_k \hat{x}_{ik}}{3} \tag{8}$$

2.3 PROMETHEE II with fuzzy weights

In the analysis multi-criteria method PROMETHEE II was also applied (Preference Ranking Organisation METHod for Enrichment Evaluations). This method belongs to the class of outranking methods. Within each criterion each couple of variants are compared; this gives the ability to determine the preferences of variants within each of the criteria. PROMETHEE II method allows you to create a ranking of variants. The procedure includes sequentially (a detailed description of the method is presented, inter alia, in [1], [14]):

- the calculation for each criterion k the distance $d_k(i, j)$ between each pair of objects i and j,
- the designation for each pair of variants of individual preference indices – at this stage, the so-called generalized criteria are used, which allow the comparison of variants within each criterion[2],
- the designation of the aggregated preference indices for each couple of variants,
- the determination of outgoing and incoming flows as well as the net outranking flow – on the basis of the net flow values the ranking is constructed (the higher value of the net flow, the higher the position in ranking).

The weights of the criteria are used at the stage of building the aggregated indices. Taking into account the fuzzy weights, at this stage triangular fuzzy numbers were used. Again, as the defuzzification method the Centre of Gravity method was used.

3 Empirical analysis

The subject of the analyses were listed companies which were part of the WIG20 index[3] in December 2017. The period of the research were the years from 2015 to 2017. The following indicators have been selected as the criteria for the evaluation of companies – which are ones of the most commonly used indicators (according to [5], [6], [11], [12]):

- return of assets ROA (net income/average total assets) – criterion 1,
- return of equity ROE (net income/shareholder equity) – criterion 2,
- P/BV indicator (price-book value) – criterion 3,
- earnings per share – criterion 4.

It was found that all criteria will be maximized. Two variants of fuzzy weights were considered. In the first, the greater suggested weight was assigned to the first two criteria (fundamental ones), in the second – the last two (market indicators). The following fuzzy weights were used:

- Variant I: $\tilde{w}_1 = (0.2, 0.3, 0.4)$, $\tilde{w}_2 = (0.15, 0.3, 0.45)$, $\tilde{w}_3 = (0.1, 0.25, 0.3)$, $\tilde{w}_4 = (0.05, 0.15, 0.25)$,
- Variant II: $\tilde{w}_1 = (0.1, 0.25, 0.3)$, $\tilde{w}_2 = (0.05, 0.15, 0.25)$, $\tilde{w}_3 = (0.2, 0.3, 0.4)$, $\tilde{w}_4 = (0.15, 0.3, 0.45)$.

The multi-criteria procedure was performed for both variants of fuzzy weights using three presented methods (in standard and fuzzy versions). The results of ordering the companies are presented in Table 1. The following markings were introduced: for the methods with standard approach three initial letters of the method name were used, for methods including fuzzy weights the name of the method was preceded by the letter F.

Co.	Variant I						Variant II					
	SAW	FSAW	TOP	FTOP	PRO	FPRO	SAW	FSAW	TOP	FTOP	PRO	FPRO
ALR	15	15	15	15	12	12	15	15	14	14	9	9
ACP	12	12	11	11	9	9	12	12	11	12	7	8
BZW	7	8	9	9	8	8	6	6	8	6	10	10
CCC	2	2	2	2	14	13	2	2	2	2	18	18

[2] After conducting preliminary analyzes, the Criterion with Linear Preference and Indifference Area was applied and the adopted thresholds of indifference and preference were 0.1 and 0.8, respectively, of the range of the criterion evaluation values.

[3] WIG20 is a stock index of the twenty largest joint-stock companies listed on the Warsaw Stock Exchange.

CPS	6	6	7	7	6	6	7	7	6	8	5	5
ENG	14	14	12	12	7	7	14	14	12	13	6	6
EUR	4	4	4	4	17	17	4	3	4	4	19	19
JSW	18	18	18	18	19	19	18	18	18	18	17	17
KGH	20	20	20	20	20	20	20	20	20	20	20	20
LTS	9	9	8	8	5	5	10	11	7	10	4	4
LPP	1	1	1	1	2	2	1	1	1	1	1	1
MBK	11	11	13	13	11	11	9	8	10	9	8	7
OPL	19	19	19	19	18	18	19	19	19	19	16	16
PEO	10	10	10	10	10	10	11	10	13	11	11	11
PGE	16	16	16	16	15	15	16	16	16	16	14	14
PGN	8	7	6	6	3	3	8	9	5	5	3	3
PKN	3	3	3	3	1	1	3	4	3	3	2	2
PKO	13	13	14	14	13	14	13	13	15	15	13	13
PZU	5	5	5	5	4	4	5	5	9	7	12	12
TPE	17	17	17	17	16	16	17	17	17	17	15	15

Table 1 Rankings of companies according to variants and methods

The analysis of the results for the variant I shows almost identical arrangement of the companies in the case of the standard and fuzzy SAW and TOPSIS methods. In addition, the correlation of the rankings of these methods is at a high level (the Spearman's rank correlation coefficient was at least 0.98). In the case of the PROMETHEE II method, the difference between the ordering of companies using the standard method and the method with fuzzy weights was also minimal (the difference of one item in the ranking concerned only CCC and PKO companies), while the rankings obtained by this method were quite significantly different from the rankings obtained by the SAW and TOPSIS methods. In the second variant, a similar situation was observed – strong correlation of rankings for the same methods in ordinary and fuzzy terms, fairly strong correlation between results obtained for SAW and TOPSIS methods and much weaker correlation of results obtained with the PROMETHEE II method compared to the other two (the rank correlation coefficient did not exceed 0.54). The largest discrepancy in results was recorded for CCC and EUR companies.

In the further stage of the research, eight-, nine- and ten-element sets were separated, which formed the basis of selecting the portfolio (companies holding the highest places in the rankings). The results of grouping are presented in Table 2.

No.	Variant (Methods) – Number of companies	Subsets of companies
1	Variant I (SAW, FSAW) – 8 Variant II (SAW) – 8 Variant II (FTOP) – 8	LPP, CCC, PKN, EUR, PZU, CPS, BZW, PGN
2	Variant I (TOP, FTOP) – 8	LPP, CCC, PKN, EUR, PZU, PGN, CPS, LTS
3	Variant I (PRO, FPRO) – 8	PKN, LPP, PGN, PZU, LTS, CPS, ENG, BZW
4	Variant I (SAW, FSAW, TOP, FTOP) – 9 Variant II (TOP) – 9	LPP, CCC, PKN, EUR, PZU, CPS, BZW, PGN, LTS
5	Variant I (PRO, FPRO) – 9	PKN, LPP, PGN, PZU, LTS, CPS, ENG, BZW, ACP
6	Variant I (SAW, FSAW, TOP, FTOP) – 10	LPP, CCC, PKN, EUR, PZU, CPS, BZW, PGN, LTS, PEO
7	Variant I (PRO, FPRO) – 10	PKN, LPP, PGN, PZU, LTS, CPS, ENG, BZW, ACP, PEO
8	Variant II (FSAW) – 8	LPP, CCC, PKN, EUR, PZU, CPS, BZW, MBK
9	Variant II (TOP) – 8	LPP, CCC, PKN, EUR, PGN, CPS, LTS, BZW
10	Variant II (PRO, FPRO) – 8	LPP, PKN, PGN, LTS, CPS, ENG, ACP, MBK
11	Variant II (FTOP) – 9	LPP, CCC, PKN, EUR, PZU, CPS, BZW, PGN, MBK
12	Variant II (PRO, FPRO) – 9	LPP, PKN, PGN, LTS, CPS, ENG, ACP, MBK, ALR

13	Variant II (SAW, TOP, FTOP) – 10	LPP, CCC, PKN, EUR, PZU, CPS, BZW, PGN, MBK, LTS
14	Variant II (FSAW) – 10	LPP, CCC, PKN, EUR, PZU, CPS, BZW, PGN, MBK, PEO
15	Variant II (PRO, FPRO) – 10	LPP, PKN, PGN, LTS, CPS, ENG, ACP, MBK, ALR, BZW

Table 2 Results of grouping companies

Designated sets differ, but these are not diametrical differences. On the basis of received subsets, portfolios were constructed based on the classic Markowitz approach (the ordinal numbers in Table 2 correspond to the portfolio number). The following optimization tasks were solved (it was assumed that the share of any share in the portfolio may not exceed 30%):

$$\begin{aligned} & S_p^2 = \sum_{i=1}^{l}\sum_{j=1}^{l} x_i x_j \operatorname{cov}_{ij} \rightarrow \min \\ & R_p \geq R_0 \\ & \sum_{i=1}^{l} x_i = 1, \quad x_i \geq 0, \quad x_i \leq 0.3, \quad i = 1, ..., l. \end{aligned} \tag{9}$$

where:

S_p^2 – variance of the portfolio,

x_i, x_j – shares of i and j shares in the portfolio,

cov_{ij} – the covariance of share return rates of the companies i and j,

R_p – the rate of return of the portfolio,

R_0 – the value of the portfolio return rate given by the decision maker[4],

l – the multiplicity of the set from which the portfolio was selected, l = 8, 9, 10.

Most of the portfolios received were of a diversified structure. It was examined what profits of the portfolios would be shaped at the end of the next three months, if on 03.01.2018, PLN 100,000 was invested in them. Table 3 presents the results obtained.

Rate of profit of portfolio compared to 03.01.2018	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**
31.01.2018	4.25	4.25	4.16	3.92	5.53	2.51	3.36	4.37
28.02.2018	-3.61	-3.61	-3.74	-4.14	0.81	-5.59	-4.22	-4.89
29.03.2018	-6.63	-6.63	-7.60	-6.42	-3.91	-8.27	-6.82	-8.40
Rate of profit of portfolio compared to 03.01.2018	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**	
31.01.2018	3.32	5.46	4.37	5.51	2.51	2.83	6.10	
28.02.2018	-5.34	-1.01	-4.89	1.10	-5.59	-5.55	1.04	
29.03.2018	-7.02	-6.04	-8.40	-3.87	-8.27	-8.20	-6.43	

Table 3 The rate of profit of portfolios (%)

The analysis of the obtained results allows to indicate the portfolios with the highest profit (the smallest loss) under the existing market conditions. The best results were recorded by P5, P12 and P15 – at the end of January they were characterized by the highest profit (not lower than 5.51%), at the end of February they were the only ones to record small profits (while other portfolios recorded losses), while at the end of March they recorded the lowest losses. The P5 portfolio was selected after using the PROMETHEE II methods (both: with standard and fuzzy weights) for the variant in which the higher weight was assigned to the fundamental criteria. The P12 and P15 portfolios were also obtained from the sets obtained after the PROMETHEE II selection, but under option II. The worst results were recorded on the P6, P13 and P14 portfolios – the lowest profits at the end of the first month (no more than 2.83%) and the largest losses at the end of the next two months. These portfolios were selected after the initial selection using the SAW, FSAW, TOP and FTOP methods (in the case of variant I for portfolio P6 and variant II for P13 and P14 respectively).

[4] The average return rate of the companies forming the basis of the portfolio was assumed.

4 Conclusion

The aim of the research was to compare the results of grouping and selection of listed companies (forming the basis for choosing a portfolio) obtained by means of selected multi-criteria methods: SAW, TOPSIS and PROMETHEE II. The analyzes were carried out, among others in a situation where the weights of the criteria are given as triangular fuzzy numbers. This approach was supposed to reflect the ambiguity of weights including the values of suggested, minimum and maximum weights. Each of the selected methods was considered in two perspectives: with suggested point weights and fuzzy weights. The analyzes carried out showed very similar rankings for the two methods included: SAW and TOPSIS, both in the standard version and with fuzzy weights (in both considered variants of weights). Similar rankings were also obtained using the PROMETHEE II method with point weights and fuzzy weights. Larger discrepancies were observed in the rankings between the results obtained by means of the first two methods and the PROMETHEE II method. Comparison of profits of portfolios selected using the Markowitz model showed that slightly better results were obtained when using the PROMETHEE II method in both versions.

At this stage of the research, for a selected group of companies, it was not possible to obtain results showing the superiority of the fuzzy approach over the standard one – in this case, the choice of multi-criteria method was the element that differentiating the results, not its version. Therefore, the research hypothesis can not be positively verified. It should be emphasized that the proposed solutions in the case of ambiguous weights require modification, and therefore the research requires continuation and another, perhaps non-standard approach.

References

[1] Brans, J.P. & Vincke, Ph. (1985). A Preference Ranking Organization Method: The PROMETHEE Method for Multiple Criteria Decision-Making. *Management Science*, *31*.
[2] Churchman, C. W. & Ackoff, R. L. (1954). An Approximate Measure of Value. *Journal of Operations Research Society of America, 2*(*1*).
[3] Jahanshahloo, G.R., Hosseinzadeh Lotfi, F. & Izadikhah, M. (2006). Extension of the TOPSIS Method for Decision-making Problems with Fuzzy Data. *Applied Mathematics and Computation*, *181*.
[4] Lai, Y.J., Liu, T.Y. & Hwang, C.L. (1994). TOPSIS for MODM. *European Journal of Operational Research*, *76*(*3*).
[5] Leszczyński, Z. (2004). *Analiza ekonomomiczno-finansowa spółki*. Warszawa: PWE.
[6] Łuniewska, M. & Tarczyński, W. (2006). *Metody wielowymiarowej analizy porównawczej na rynku kapitałowym*. Warszawa: PWN.
[7] Pośpiech, E. (2017). Multi-criteria fuzzy modelling in the issue of portfolio selection. In M. Čulik (Ed.) *Proceedings (Part III.) of the 11th International Scientific Conference on Financial Management of Firms and Financial Institutions* (pp. 700–707). Ostrava: VŠB-TU Ostrava.
[8] Pośpiech, E. (2018). Rozmyte modelowanie we wspomaganiu decyzji inwestycyjnych. *Organizacja i Zarządzanie*, *113*, 375–384.
[9] Pośpiech, E. & Mastalerz-Kodzis, A. (2015). Wybór metody wielokryterialnej do wspomagania decyzji inwestycyjnych. *Organizacja i Zarządzanie*, *86*, 379–388.
[10] Pośpiech, E. & Mastalerz-Kodzis, A. (2016). Zastosowanie metody TOPSIS w ujęciu rozmytym do selekcji walorów giełdowych. *Organizacja i Zarządzanie*, *96*, 395–404.
[11] Tarczyński, W. (2001). *Rynki kapitałowe. Metody ilościowe*. Warszawa: Polska Agencja Wydawnicza PLACET.
[12] Tarczyński, W. (2002). *Fundamentalny portfel papierów wartościowych*. Warszawa: PWE.
[13] Trzaskalik, T. (Ed.) (2006). *Metody wielokryterialne na polskim rynku finansowym*. Warszawa: PWE.
[14] Trzaskalik, T. (Ed.) (2014). *Wielokryterialne wspomaganie decyzji*. Warszawa: PWE.
[15] Tzeng, G. H. & Huang, J. J. (2011). *Multiple Attribute Decision Making. Methods and Applications*. London: CRC Press.
Data: https://www.bankier.pl/ (23.04.2018).
Data: https://www.gpw.pl (23.04.2018).
Data: infostrefa.com/infostrefa/pl/archiwum (25.04.2018).

Application of game theory in e-commerce

Alena Pozdílková[1], Ondřej Komárek[2]

Abstract. The application of principles of game theory to business and economics is as old as the mathematical discipline itself. Reasons for this are quite clear: Interactions with customers, business partners, and competitors, in most cases, closely match models in game theory. This is further enhanced by the expansion of information technology and e-commerce. Here the customer and trader get into formal situations that can be modeled similarly to rational players of certain games. Game theory has recently begun to address increasingly large Czech and foreign online stores. Almost every e-shop has a large amount of well-sorted data about its customers, their behavior and price reflection on product sales in the form of past orders. Nevertheless, only a few of the possible uses have been explored, mainly in the field of intelligent product recommendation to individual customers, according to extensive statistics. But these options do not end here. The application of the game theory is also possible, in combination with statistical methods, to create complex strategic procedures to maximize profit. This can be achieved, whether by reordering products by customer groups, individual flash sales or targeted price manipulation. The aim of this article is to apply game theory and statistical methods (silhouette analysis, cluster analysis etc.) to the e-commerce sphere. This will create a theoretical basis to help with practical calculations. In addition to the theoretical part, the whole issue will be tested using an automated decision support tool in a practical online store.

Keywords: Game theory, e-commerce, silhouette analysis, cluster analysis.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

In the last decade there has been an enormous increase in e-commerce. Today almost every small retailer or extensive shopping chain uses its online shop in some form. This is the dream business. With costs substantially negligible, as opposed to a stone store, it is possible to reach out to customers around the world. At the same time, moreover, the cost of entering this business sphere has declined significantly. The customer is accessible through an online shop, which can be created on the basis of a simple template which costs only a few thousand crowns per month, or as a complex, custom-made software whose price is in the order of millions of crowns. That's why sellers try to distinguish themselves in different ways. These include low prices, a variety of offered services, and individual customer access.

The aim of this work is the complex processing of the theory of games, and their principles. These, along with selected statistical techniques, will be used to understand the customer, and anticipate future decision making.

1.1 Methodology

This chapter will specify individual entities which will be used in this work. The game in the field of e-commerce is best defined as the endlessly repeating Bayesian game, through cooperative negotiation. Endless repetition of the game is the main focus of the seller, which will put long-term income ahead of short term goals. For the customer, this game is also endless by definition. [3]

Player

The player will think of such an entity that will either endlessly participate in the game or periodically participate in any of its rounds. A set of all current and past players will be labeled N. The player will be marked as $n, n \in N$.

Customer

[1] University of Pardubice, Faculty of Electric Engineering and Informatics, Department of Mathematics and Physics, Studentská 95, 532 10 Pardubice 2, Czech Republic, e-mail: alena.pozdilkova@upce.cz
[2] University of Pardubice, Faculty of Electric Engineering and Informatics, Department of Mathematics and Physics, Studentská 95, 532 10 Pardubice 2, Czech Republic

The customer will be the player whose payout and the result of cooperative bargaining will be the seller's product. The set of all players will be denoted as C and individual player as c with the fact that $c_i \in C$ and $C \subseteq N$ where $i \in \{1, \dots, n_c\}$.

Customer history h_i, where $h_i \in H$ is the set of all moves that the customer has already made in the game. History is known only to vendors d, and to the customer c_i to whom this history belongs.

Customer parameters pa_{c_i} are a set of values that identify it. It is not excluded that $pa_{c_i} = pa_{c_j}$, for $i = j$ may occur. The customer group G is characterized by the parameters of all customers who belong to it.

Vendor

The vendor will be a type of player who will participate in a cooperative bargaining with the customer, and his payout will be earnings from the product the customer has purchased. Each dealer will be denoted by the letter d and the set of sellers with the letter D. Each $d \in D, D \subseteq N$ and $d_i = c_i$ for all $i \in N$, where $i \in \{1, \dots, n_v\}$.

Customer type Θ_c, type of dealer Θ_d or simplified type of player Θ_i will be such information about the previous course of the game, which will be private for the given player.

Product

The product will also be understood in this work as a negotiation item. The 1st product will be marked as p_i. It is true that $p_i \in P$, where $P = \{p_i\}$ is the set of all products in the game, for $i \in \{1, \dots, n_p\}$. Each product will be priced at. The product will be taken as a customer payment c.

The price of the product p_i is the amount determined by vendor d, it is applied to the product for a long time and fulfills $pr_i \in PR$, where PR is the set of all prices. This price will also reflect payee sales. A special case opr_{ij} is to stop the price of the product p_i at the time of purchase o_j.

The discount for ds_i of product p_i is the amount that is applied to the product strictly for a short period of time t, it can also be applied only to a certain group of customers g or even to an individual customer c. For each product p_i, where $i \in \{1, \dots, n_p\}$: $pr_i - ds_j \in PR$. A special case ods_{ij} is the discount of product p_i at the time of purchase o_j.

The expiration of product p_i is a non-negative integer determining the trader's estimate for how long a customer will want to restore the product. Expiration is very important for non-consumable goods that the customer will most likely not buy repeatedly.

The addendum $pc_{ij} \in PC_j$ is such a product p_i, which the merchant regards as a suitable complement of the product p_j. For example, table to chair. Product p_j may have multiple add-ons.

Purchasing

Purchasing or_i, where $or_i \in OR$ is represented by a set $or_i = \{op, opr, opa, ods, c, pps\}$, where op means products, opr their purchase number, opa their prices, ods their discounts, c customer and pps purchasing parameters. At the same time, $or_i \in H$, even if the purchase is not completed and is included in the calculation of the strategy.

Purchase parameters pps_{ij} are the customer's parameters at the time of purchase.

Customer group

The customer group g_i is such a set of players that $g_i \subseteq C$. The customer group is distinguished by the headquarters, which has a common property as well. For example, geographical location, time, or date of purchase. Thanks to the fact that the history of this group h_i is available, it is possible to predict the customer's behavior without having its own history.

2 Splitting customers into groups

As mentioned in the previous chapter, in the vast majority of cases, it is not possible to create an acceptable strategy for analyzing the history of only one customer or player. This is mainly due to the insufficient data set. In these cases, it is ideal to divide players into larger or smaller groups based on common features. The size of the group will be given by the foremost level of significance.

2.1 Cluster analysis

Because of the unknown number of parameters, cluster analysis appears to be the ideal method of group formation. It analyzes groups of objects based on information found in the data describing the objects or their relationships.

The goal is that the objects in the group will be similar or related, and distinct from objects in other groups. The higher the homogeneity in the group and the bigger the difference between the groups, the better the clustering.

2.2 Selection of suitable variables

The first step before applying clustering algorithms is to select appropriate variables to compare customers. Clustering using the address of the customer is also advantageous, whether due to economic or climatic conditions. The bigger problem in this case is that it is necessary to use quantitative units for the purpose of inclusion, but these are not so easy to obtain. Most of the registered customers are known by their IP address and by the same number of registered customers, the name of the country and the place of residence. Therefore, it will be necessary to use a mechanism that will be able to convert IP addresses or residence to latitude and longitude. In this case, it will be necessary to determine how fine the granulation will be ranked by the customer. Of course, the most appropriate would be to determine the location according to the city or village where the customer resides. But this would necessitate a large database with all the world's housing estates or the use of external services. An external service would in turn cause unnecessary delays in calculations, and at the same time its outage would significantly reduce the accuracy of the entire order. Therefore, the most appropriate combination of placement and place of residence will be the best.

The other unit that seems to be very convenient is the time and date in which the customer is doing business. These can be explained by the fact that they will be very likely to buy something in the morning or in the afternoon at the same time as the holidaymaker. In most cases, the schoolteacher will not buy late in the evening. The most appropriate will be the first quantitative unit of arrival 1-12, day of week 1-7 and then hour 0-24. Although these values will be shared across most online stores, it may be possible for specific implementations to expand. For example, there may be shops, some of which are required by the customer, so that there may be a number of age-related customers depending on the age group, and that is why they can adapt the products offered. If a customer registers, it will be necessary to create a weighted average of his statistics and to handle outdated hotspots because the customer who shops at an unusual time, is likely to be only an isolated case that could influence the calculations for a small amount of data.

2.3 Cleaning parameters

With repeated visits to the online store, it is possible to place a specific customer more accurately. But as with most statistical files, it is necessary to clean the data before processing. Due to the nature of the data, it is not necessary to look at the missing values when cleaning the parameters. On the other hand, the existence of extreme values is more than likely. A customer shopping around noon will remember before Christmas that he forgot to buy a gift and order at midnight. Or is on vacation and orders from another country.

purchase	month	day	hour	longitude	latitude
pps_1	4	5	11	49.652456	16.259766
pps_2	6	6	12	49.652456	16.259766
pps_3	8	1	9	49.652456	16.259766
pps_4	10	2	13	35.320802	25.138551
pps_5	11	6	10	49.652456	16.259766
pps_6	12	1	23	49.652456	16.259766

Table 1 Customer purchases with extremes before cleaning

For circular values such as the hour or the month of the year, it is necessary first to transform into a point in 2d space using

$$x_x = \sin\frac{2\pi x}{z}, x_y = \cos\frac{2\pi x}{z} \tag{1}$$

where $1 \leq x \leq z$ [4], otherwise there would be errors in further calculations because 1 hour in the morning is closer to 23 hours in the evening even if it is not evaluated.

purchase	month-x	month-y	day-x	day-y	hour-x	hour-y	longitude	latitude
pps_1	0.866	-0.5	-0,975	-0.223	0.259	-0.966	49.652456	16.259766
pps_2	0	-1	-0.782	0.623	0	-1	49.652456	16.259766

pps_3	-0.866	-0.5	0.782	0.623	0.707	-0.707	49.652456	16.259766
pps_4	-0.866	0.5	0,975	-0.223	-0.259	-0.966	35.320802	25.138551
pps_5	-0.5	0.866	-0.782	0.623	0.5	-0.866	49.652456	16.259766
pps_6	0	1	0.782	0.623	-0.259	0.966	49.652456	16.259766

Table 2 Customer purchases with extremes after cleaning

Remote modified z-score will be used to detect outliers. This approach modifies the standard z-score so that it is not affected by high extremes. Modified z-score M_i for x_i will be calculated as follows:

$$M_i = \frac{0.6745(x_i - \tilde{x})}{MAD} \tag{2}$$

where $\tilde{x}$ is the median of the data file, and $MAD = median\{|x_i - \tilde{x}|\}$. In [5] is recommended that observation be labeled as a distant value if $|M_i| > 3.5$. The problem with this method occurs when more than 50% of observations have the same value, in which case MAD is equal to 0 and the equation fails.

In this case, IBM [6] recommends replacing the Median absolute deviation (MAD) with $MeanAD$ (Mean absolute deviation), where

$$MeanAD = \frac{\sum|x_i - \tilde{x}|}{n} \tag{3}$$

Due to the nature of the data, it is not appropriate to cut off these outlying values. All values will therefore be evaluated by weight 1, which in the case of remote values will be adjusted by Hubert's weighting function, which is calculated as follows:

$$W_c(x) = \min\left\{1, \frac{c}{|x - \hat{\mu}|}\right\} \tag{4}$$

where c is the bending criterion, in this case, c = 1.28 will be used, which gives 90% asymptotic efficiency and is less prone to a data set contaminated by remote values. $\hat{\mu}$ represents the location estimate, which in this case is equal to the absolute value of the median $|\tilde{x}|$ for $|\tilde{x}| \neq -x$. and $|\tilde{x}| + 1$ for $|\tilde{x}| = -x$. [7] It follows from this equation that the more the value differs from the median, the less weight will be assigned to it. Then the weighted average is derived from the values and the final customer parameters are obtained.

month-x	month-y	day-x	day-y	hour-x	hour-y	longitude	latitude
-0.228	0.061	0	0.341	0.158	-0.590	49.401	16.509

Table 3 Cleaned customer parameters

2.4 Creating clusters and optimizing number of clusters

Due to the complexity of generating adequate clusters for large data sets, it would be best not to create clusters dynamically based on new arrivals, but always at a time interval that would include all entities that were included in the system since the last recalculation. In the vast majority of cases, rebuilding will take place at the time of the smallest Internet business activity. Therefore, the transition to a standard weekly interval will be performed only after the minimum size of the data file, specified in the implementation, has been reached. The standard interval will be, at most, once a week. But, in the case of abnormally large data growth, it is possible to check before this standard interval, but no more than once a day. This abnormal increase will be determined by the percentage constant, depending on the implementation. If this constant is selected as 10%, it will mean that if the increase of the previous day is 10% higher than the average, the premature conversion of clusters will be triggered. This diameter will, of course, be adjusted by 5% of the highest and lowest values. Given that the dataset will always be of a quantitative nature, the K-means ++ algorithm will be used as a clustering method.

Of course, in the case of dynamically growing data, it is not advisable to be static. Therefore, in this case, only the initial minimum number of clusters will be represented. This will continue to be edited using silhouette analysis.

The optimum number of clusters is determined by the Silhouette index in the silhouette analysis. At first, an initial count of K_s clusters is set during implementation. The K_s will only estimate the initial number of clusters, and in most cases, $K_s = 1$ will be absolutely sufficient. Only for longer-lasting online stores with a large number of customers, it is advisable to specify a higher value of K_s and thereby reduce the computational difficulty of the

primary cluster analysis. The definition of the silhouette index is based on the Rousseeuw silhouettes [2], [8], which are designed as a graphic aid indicating how well each object is classified from the cluster, see [1].

3 Determination of the strategy

If the customer history is already known, or the customer's predicted history, it is possible to determine its probable strategy. The main goal of a customer's strategy will be to satisfy its need and to spend the least amount of money. On the other hand, the dealer's goal is to maximize profit. It is likely that these two strategies will vary considerably in many cases. The following chapters will describe how to create a customer and reseller strategy. Then, the main goal will be to find Nash's equilibrium strategies between the two sets so as to maximize profits while meeting customer needs.

3.1 Customer strategy

Due to the knowledge of all previous customer decisions, it is possible to determine its probable strategy. This strategy will, of course, change during the customer's stay in the online store. For example, if a customer puts a p_1 product in the basket, it will be more likely that his strategy will match one of the previous purchases in which the product was selected. The alignment of products in a given strategy will be accomplished using scales that will vary based on the situations outlined above. Implementation will determine the multiplier to give preference to customer history, if any, before the history of its group. At the same time, however, if the customer does not have any product history, but his cluster does, only the history of the cluster is taken. [2]

Each product also has a defined expiration time. It is responsible for taking into account the fact that not every item has a consumer character. However, the exclusion of unexposed products from the strategy is not appropriate. Therefore, the unexposed product will be treated as a product purchased by a customer group. This means that the product that was purchased before it expires, will not be multiplied by the x factor when calculating the purchase. At the same time, however, the benefits of this product will be favored. The advantage will be that the values calculated for $p_j \in PC_i$ will be added to the given purchase:

$$opa_{jnew} = opa_j + opa_i \max\left(1, \frac{x}{2}\right) \tag{5}$$

where $i, j \in \{1, \dots n\}, i \neq j$.

Product prices at the time of purchase are used as weights for calculating the average. Weights w must be set so that preference is given to those product purchase values where the price is as close to the current price as possible. Because with a lower price it is more likely that a customer purchases a product, it is also necessary to prefer those purchases where the price was higher than the current price. For w_i, where $w_i \in W_j$ had such a predictive value, it is necessary to convert the price opr and its normalization based on the following formula:

$$w_i = \frac{1}{1 + \left(\widetilde{pr_i} - \widetilde{opr}_{ij}\right)} \tag{6}$$

where $\widetilde{pr_i} = \frac{pr_i}{\|y\|}, \widetilde{opr}_{ij} = \frac{opr_{ij}}{\|y\|}$ and $\|y\| = \sqrt{\sum_{j=1}^{n} opr_{ij}^2 + pr_i^2}$ for such products p_i, in the order or_j for which $opa_{ij} > 0$ for $i, j \in \{1, \dots n\}$.

The last step is to calculate the weighted average value of purchases, where the prices of products will serve as scales, so that there is just one value for each product currently on sale. Thanks to weightings, the price of the product at the time of purchase will also be included in the calculation.

$$P(pr_i) = \sum_{j=1}^{n} w_i . opa_{ij} \tag{7}$$

3.2 Vendor strategy

Vendor strategy will be more straightforward than the customer's strategy. The main indicator will be what payout the sale of the product offers. This indicator will be the Product Payoff Coefficient, PPC. This will be a non-negative integer, the amount of which will correspond to the payout when selling the product. In many cases, it will be possible to make the product more attractive to the customer without significantly reducing its PPC value. The primary goal of the vendor's strategy will be to create Nash's balance so that the PPC is as high as possible. Unlike a customer's strategy, the vendor's strategy does not change during the purchase. [2]

3.3 Equilibrium strategy

To calculate Nash equilibrium, it is first necessary to place customer and vendor strategies into the following matrix:

Customer Vendor	0.319	0.386	0.156	0.139
0.383	**0.383; 0.319**	0.383; 0.386	0.383; 0.156	0.383; 0.139
0.115	0.115; 0.319	**0.115; 0.386**	0.115; 0.156	0.115; 0.139
0.137	0.137; 0.319	0.137; 0.386	**0.137; 0.156**	0.137; 0.139
0.365	0.365; 0.319	0.365; 0.386	0.365; 0.156	**0.365; 0.139**

Table 4 Nash equilibrium

It has to be taken into account that whatever the vendor's strategy is, the customer always chooses the result. Therefore, it is possible to work only with values on the diagonal of the table. In this case, the customer's strategy would have been to buy the product p_2, followed by p_1.

It is obvious, however, that such a result would be disadvantageous for dealers in terms of their own strategy. Therefore, by multiplying the value of customer strategy and seller strategy in the table, new values are obtained. These will serve as a metric that, with respect to the vendor's strategy, will determine the order of the products in which it is best to sell to the customer, so that the predicted strategy of the customer is maintained. In this case, the table would look like this:

Product	p_1	p_2	p_3	p_4
Weight	0.122	0.044	0.021	0.051

Table 5 Nash equilibrium, weights of products

Interpreting this table then depends on the implementation. In most cases, it can be used as a guideline for the aggressiveness of the product to the customer. This can be a preference for a search or a preferred position on the opening page. Of course, it is not necessary to use only conservative passive methods. Although a vendor cannot directly influence the customer's strategy based on his or her strategy, he can change the game rules themselves. The main tool for these changes is the price of the product.

4 Conclusions

The aim of this article was to apply game theory and statistical methods (silhouette analysis, cluster analysis etc.) to the e-commerce sphere. Very useful results were founded. All of these procedures will be implemented and tested in a real online store with a large database of customers and after testing it will be used in a real environment.

Acknowledgements

The contribution has been supported by University of Pardubice.

References

[1] Vipin, K. (2005). *An Introduction to Cluster Analysis for Data Mining.* Addison: Wesley. ISBN 978-0321321367
[2] Rousseeuw, P. J. (1986). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Fribourg: University of Fribourg.
[3] Maňas, M. (2001). *Teorie her a optimální rozhodování*. Praha: SNTL. ISBN 80-7226-385-4
[4] Nicholas J. Cox (2006) Speaking Stata: *In praise of trigonometric predictors*. StataCorp LP: 2006. ISBN 978-0321321367
[5] Songwon Seo (2002) *A Review and Comparison of Methods for Detecting Outliersin Univariate Data Sets*. BS, Kyunghee University: 2002.
[6] IBM Knowledge Center Modified z score [online]. IBM [cit.2018-04-16]. URL: https://www.ibm.com/support/knowledgecenter/en/SSWLVY_1.0.0/com.ibm. spss.analyticcatalyst.help/analytic_catalyst/modified_z.html
[7] Ken A. Aho (2007) Foundational and Applied Statistics for Biologists Using R CRC Press: 2007. ISBN 978-1439873397
[8] P. J. Rousseeuw (1986) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. University of Fribourg: 1986.

Efficiency of agricultural farms measured by data depth and DEA: comparison and empirical study

Miroslav Rada[1], Ondřej Sokol[2], Michal Černý[3]

Abstract. We discuss seemingly independent concepts that can be utilized for measuring the productive efficiency of decision-making units (or, geometrically, multivariate datapoints): classical DEA models on one side and fixed-orthant halfspace depth on the other side.
Halfspace depth, introduced by Tukey, is a natural generalization of the concept of median to multivariate data. To serve the purpose of measuring efficiency of decision-making units, its definition has to be altered to measure depth in one fixed orthant depending on structure of input. Although this alteration is a simplification from computational point of view, the computation remains time-consuming.
Hence, we set the polynomially-computable DEA that provides us with real-valued efficiency against superpolynomial halfspace depth, which finds equivalency classes of efficiency, at best. Despite these drawbacks, we show that halfspace depth provides more information about the decision-making units, in particular about indispensableness of individual units.
These findings are justified by empirical study on FADN data of agricultural farms provided by FADN; the experiments also uncover resemblance between both the efficiency measures.

Keywords: agricultural farms, DEA models, efficiency, data-depth.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Topic and goal of the paper. This paper compares one common variant of data envelopment analysis with an (appropriately modified) statistical method for ranking data points – Tukey's halfspace depth. The goal is to point to the similarities and differences, show that modified halfspace depth provides a more robust information, and, finally, demonstrate the methods on real world data from agricultural industry.

Structure of the paper. In Section 1, we describe the ideas and basics of DEA and depth measures, define them and introduce the notation of input data and our setup. Section 2 shows some properties of the methods and discusses the complexity of computation. Section 3 presents the design of our computational experiments and finalizes the paper with their results.

Our setup and notation of units. Thorough the paper, we are given a list $\mathcal{U}$ of p units $U_1, \ldots, U_p$. Every unit, say U_k, is associated with $m + n$ given parameters. For simplicity, we write $U_k = (i_{k1}, \ldots i_{km}, o_{k1}, \ldots, o_{kn})^T$. The first m parameters for every unit are conveniently called *inputs*, while the other n are *outputs*. We assume that the inputs are to be minimized and the outputs to be maximized. Let $q := m + n$ be the total number of variables.

To avoid degenerate cases, we also assume that $p > q$ and that no $q + 1$ units are affinely dependent. This allows to focus on ideas rather than on technicalities, furthermore, we premise that this assumption is valid for the data we used in Section 3 for computational experiments.

1.1 Data envelopment analysis

Since the very first introduction of *data envelopment analysis* (hereinafter *DEA*) by Charnes et al. in 1978 [5], DEA models proved to be useful tools for comparing efficiencies of homogeneous processing *units*. Units are understood as blackboxes converting some *inputs* to some *outputs*. The values of inputs and outputs are the only data to work with. Since there can be multiple inputs and outputs, one has to aggregate them somehow.

[1]University of Economics, Prague, Department of Financial Accounting and Auditing, W. Churchill sq. 4, Prague 3, 130 67, Czech Republic, miroslav.rada@vse.cz
[2]University of Economics, Prague, Department of Econometrics, W. Churchill sq. 4, Prague 3, 130 67, Czech Republic, ondrej.sokol@vse.cz
[3]University of Economics, Prague, Department of Econometrics, W. Churchill sq. 4, Prague 3, 130 67, Czech Republic, cernym@vse.cz

DEA uses a quite simple aggregation tool: a *linear scalarization*. This simplicity brings several advantages: firstly, the obtained results have relatively simple interpretation and the idea is relatively easy to grasp, secondly, the *geometry* of the approach is "flat", allowing for quite simple visualizations and better understanding of the problem, and thirdly, the methods used for implementation of the problem are well-known and feasible from the computational-complexity point of view.

DEA does the following when computing efficiency of a given unit U_k: It takes the set of all units available and tries to combine them into the best fictional unit U_k^* that *dominates* U_k in Pareto sense; this is actually done by fixing inputs or outputs of U_k to some level and rescaling its outputs or inputs. If such a best fictional unit is U_k itself, then U_k is efficient, otherwise one can compute its relative efficiency against the U_k^*. Geometrically, DEA tries to find the best fictional unit on the boundary of *convex hull* of all units. Such a convex hull could be called *data envelopment* — this is where the name of DEA comes from.

Our DEA variant. There are many variants of DEA, which differ in several aspects. The most common aspects are the *returns-to-scale* setting (*constant* or *variable*) and *orientation* of analysis (*input-oriented* or *output-oriented*). The former aspect distinguishes whether parameters of a unit can be scaled arbitrarily or not. Here, we focus on variable returns-to-scale [2], since it is more suitable for our case of agricultural industry. The latter aspect does not influence obtained results in the sense of amount of information gained. The efficiency in an output-oriented model says, roughly spoken, how many times the output parameters should be increased to make an unit efficient, while in the input-oriented model, it provides similar information about reducing of inputs. Here, we focus on input-oriented model, since it is more suitable for our dataset.

For a given unit U_k, the linear program (P_k) can be used for computation of DEA efficiency:

$$\begin{aligned}
&\text{minimize} && \theta =: E_k && && \\
&\text{subject to} && i_{kr}\theta \geq \sum_{\ell=1}^{p} x_\ell i_{\ell r}, && r = 1,\dots,m && \\
& && o_{kj} \leq \sum_{\ell=1}^{p} x_\ell o_{\ell j}, && j = 1,\dots,n && (\mathrm{P}_k)\\
& && \sum_{\ell=1}^{p} x_\ell = 1, && && \\
&\text{by changing} && \theta, x_\ell \geq 0 && \ell = 1,\dots,p,. &&
\end{aligned}$$

Definition 1 (DEA efficiency)**.** The optimal value E_k of (P_k) for a given unit U_k is called *efficiency of* U_k. A given unit U_k is *efficient (in the DEA sense)*, if its efficiency is equal to 1. Otherwise, it is *inefficient (in the DEA sense)*.

Information to be gained from DEA. DEA provides various types of information: usually, it is utilized to distinguish efficient and inefficient units. It is sometimes also used for *ranking* units. However, for this purpose, DEA itself is not quite robust, since the ranks can be dependent on shifting (and in some cases also on scaling) of variables. One of the goals of this paper is to discuss some methods that are based on similar concept of efficiency while producing ranking with more desirable properties.

Also, DEA gives the information about how all the inputs (or outputs) of an inefficient units shall be simultaneously scaled to obtain an efficient unit. This is a praised property, however, it shall be recalled that the efficiency of an inefficient unit depends on the initial setup (i.e. shifting and scaling of input and output variables).

1.2 Data depth measures

Halfspace depth and its relation to DEA. In 1975, Tukey [16] proposed so-called *halfspace median* or *Tukey's median* as a multivariate generalization of the usual univariate median. The main purpose is to somehow measure "centrality" or "being inside" for a given point in a dataset.

In the univariate case, a median is a point that is the deepest in the data set – if one wants to leave the data set in either direction from the median, he meets at least one half of points of the dataset. Tukey's idea was quite similar. Halfspace depth of a given point U with respect to dataset $\mathcal{U} = \{U_1, \dots, U_p \in \mathbb{R}^q\}$ is the minimal number of points from $\mathcal{U}$ contained in a halfspace H containing U. The normal vector of the halfspace H can be viewed as the direction of leaving the dataset; a point in the dataset "is met" if it belongs to the halfspace H. The less points, the shallower U is, the more points, the deeper U is. In the univariate case, there are only two possible normals (towards to $-\infty$ and towards to $+\infty$), in the multivariate case, the best normal from the whole $\mathbb{R}^q$ has to be taken into account.

This is actually quite similar to DEA. DEA measures "distance" in the convex hull of the whole dataset in a direction that maximizes outputs and minimizes inputs. An analogy offers here: let's restrict halfspace depth (or any other suitable data depth) such that it will measure depth only in the suitable orthant, and consider depth as the "distance".

The definition of halfspace depth follows. The notation is used to match the notion of units and their inputs and outputs. Definition 2 is the standard definition of halfspace depth, Definition 3 is the definition which measures the depth only in a specified orthant.

Definition 2 (Halfspace depth)**.** Consider a dataset $\mathcal{U} = \{U_1, \ldots, U_p \in \mathbb{R}^q\}$. A number

$$\min_{x \in \mathbb{R}^q} |\{U_\ell \in \mathcal{U} \mid U_\ell^T x \geq U_k^T x\}| \tag{1}$$

is called *halfspace (or Tukey's) depth of* U_k.

Notation intermezzo. For a real vector a, the symbol $\mathrm{diag}(a)$ denotes diagonal matrix of a. The symbol $\mathrm{conv}(S)$ denotes the convex hull of S. For two sets $S_1, S_2 \subseteq \mathbb{R}^q$, the sum is defined as follows: $S_1 + S_2 \coloneqq \{s_1 + s_2 \mid s_1 \in S_1,\ s_2 \in S_2\}$. Let $S^q \coloneqq \{-1, 1\}^q$ be the space of sign vectors of length q. An orthant $\mathcal{O}_s \coloneqq \{\xi \in \mathbb{R}^q | \, \mathrm{diag}(s)\xi \geq 0\}$ can be associated to every $s \in S^q$.

Definition 3 (Orthant-specific halfspace depth)**.** Consider a sign vector $s \in S^q$ and a dataset $\mathcal{U} = \{U_1, \ldots, U_p \in \mathbb{R}^q\}$. A number

$$T_k^s \coloneqq \min_{x \in \mathcal{O}_s} |\{U_\ell \in \mathcal{U} \mid U_\ell^T x \geq U_k^T x\}| \tag{2}$$

is called *halfspace (or Tukey's) depth of* U_k *in orthant* s.

Other concepts. Through the decades from the Tukey's initial concept, many various ideas were proposed, e.g. Oja's depth, simplicial depth or regression depth. We refer to quite recent surveys on data depth [12, 15]. All the concepts possess some advantages or disadvantages. Among them, halfspace depth together with regression depth exhibits desirable properties, such as *affine equivariance*, *vanishing at infinity*, *monotonicity* and *robustness*, see [6, 14, 17]. This, together with its suitable geometric interpretation, makes halfspace depth the most suitable for our purposes.

Convex hull peeling depth. Aside from halfspace depth, so-called *convex hull peeling depth* is also of interest, since it is closely interconnected with DEA models. Convex hull peeling depth was introduced by Barnett [3].

Definition 4 (Convex hull peeling depth)**.** Assume a dataset $\mathcal{U} = \{U_1, \ldots, U_p \in \mathbb{R}^q\}$.
(a) We define *jth convex hull layer*, denoted by L_j, as

$$L_j \coloneqq \{U_\ell \in \mathcal{U} \mid U_\ell \text{ is an extremal point of } \mathrm{conv}(\mathcal{U} \setminus \cup_{k=1}^{i-1} L_k)\}. \tag{3}$$

(b) A *convex hull peeling depth* of a point U_k is the number j of the convex hull layer L_j satisifying $U_k \in L_j$.

Note that every data point is associated to a layer. The DEA-alike modification of convex hull peeling depth follows. The convex hull is now sought only in one orthant. This is ensured by adding the orthant $\mathcal{O}_{-s}$ to the convex hull of units.

Definition 5 (Orthant-specific convex hull peeling depth)**.** Consider a sign vector $s \in S^q$ and a dataset $\mathcal{U} = \{U_1, \ldots, U_p \in \mathbb{R}^q\}$.
(a) We define *jth orthant-specific convex hull layer*, denoted by L_j^s, as

$$L_j^s \coloneqq \{U_\ell \in \mathcal{U} \mid U_\ell \text{ is an extremal point of the set } H_j^s \coloneqq (\mathrm{conv}(\mathcal{U} \setminus \cup_{k=1}^{j-1} L_k^s) + \mathcal{O}_{-s})\}. \tag{4}$$

(b) A *convex hull peeling depth in orthant* s of a point U_k, denoted by P_k^s, is the number j of the orthant-specific convex hull layer L_j^s satisifying $U_k \in L_j^s$.

2 Properties and computation of orthant-specific halfspace depth

Here, we examine several properties of both the forementioned depth measures and discuss the complexity of their computation.

We recall our notation: U_k is q-dimensional vector, $q = n + m$, first m elements are inputs, the last n elements are outputs. Since our DEA setup defines one orthant that is of our interest, we set

$$s^* \coloneqq (\overbrace{-1, \ldots, -1}^{m}, \overbrace{1, \ldots, 1}^{n}).$$

From now on, we will write $T_k \coloneqq T_k^{s^*}$ and $P_k \coloneqq P_k^{s^*}$. Also, we repeat the assumption that $p > q$ and no $d + 1$ units are affinely dependent.

Lemma 1 (Equivalence of efficiency)**.** *The following statements about a unit U_k are equivalent:*
(a) U_k is efficient in the DEA-sense, i.e. $E_k = 1$,
(b) U_k has orthant-specific halfspace depth $T_k = 1$,
(c) U_k has orthant-specific convex hull peeling depth $P_k = 1$.

Proof (sketch).
- (a) $\Rightarrow$ (b). If U_k is efficient, then the program (P_k) has an optimal solution (x^*, θ^*) with $\theta^* = 1$. This means that it is not possible to build a fictional unit from a convex combination of units in $\mathcal{U}$, that is strictly better than the unit U_k in all the variables. Since no $d+1$ units are affinely dependent, U_k has to be a vertex of convex hull of $\mathcal{U}$, hence, its depth is 1.
- (b) $\Rightarrow$ (c). Denote by x^* the minimizer in (2). Such an x^* witnesses that U_k is a vertex of $H_1^{s^*}$ from (4).
- (c) $\Rightarrow$ (a). If U_k is a vertex of $H_1^{s^*}$, it means that (P_k) is feasible with $\theta = 1$ and x set to kth unit vector. This means that for these values, it is optimal and U_k is efficient in the DEA sense. □

Lemma 2 (Interpretation of T_k)**.** *The number $t = T_k - 1$ is the least number of units to be removed for U_k to become efficient in the DEA sense.*

Proof. Assume x^* is minimizer in (2). Then, after removing units from the set $\{U_\ell \in \mathcal{U} \mid U_\ell^T x^* > U_k^T x^*\}$, the vector x^* witnesses that the new depth is 1.

On the other hand, removing less than t units is not sufficient, since the halfspace depth does not decrease to 1. □

Complexity of computation of P_k. Lemma 1 offers a way to compute peeling depths. One can run DEA analysis to identify $L_1^{s^*}$. Then, P_k of all the efficient units can be set to 1 and these units can be removed. This can be iterated for the second and every other convex-hull layer.

At least one unit does have to be in every layer. This means that there are at most p layers. In every layer, all the remaining units must be tested for efficency. This gives an upper bound on number of linear programs of form (P_k) to be solved: $p(p-1)/2$.

Complexity of computation of T_k. By transformation to densest hemisphere problem ([11]), halfspace depth is know to be NP-comlete problem. The restriction to one orthant does not help much. Nevertheless, there are several algorithms to compute it. We are especially interested in the algorithms that compute halfspace depth by enumeration of cells of a hyperplane arrangement. Such a method was probably firstly mentioned by Eppstein [9]. The idea is the following. Consider (1). Assume we are computing halfspace depth of U_k. Consider a unit U_ℓ. The hyperplane $\{\xi \mid U_k^T \xi = U_\ell^T \xi\}$ dissects $\mathbb{R}^q$ to two halfspaces. In one of them points contribute to halfspace depth of U_k, in the other they don't. Hence, such hyperplanes dissect the space of all possible minimizers of (1) (i.e. $\mathbb{R}^q$) to (convex) polyhedral regions with the same halfspace depth, so-called *cells*. Since we have no $d+1$ affinely dependent units, all the regions of interest are fulldimensional. By [4], the number of fulldimensional cells of a hyperplane arrangement in $\mathbb{R}^q$ of p hyperplanes is $O(p^q)$. However, these cells can be enumerated *as a stream* using output-sensitive algorithms [1, 13], which enables to employ promising heuristics that could help in some cases.

To adapt the algorithm of [13] for orthant-specific depth, it is actually sufficient to restrict the enumeration to cells which have nonempty intersection with $\mathcal{O}_{s^*}$. This is what we did in our implementation used for experiments in Section 3. For more details about the original algorithm, see [13].

3 Computational experiments

We compared all the above ranking methods on real world agricultural data. The description of the data and results itself follow. Actually, the design of the experiment follows up and builds on the conference talk [10].

3.1 Data

Data source. The data comes from *Farm Accountancy Data Network* (FADN) [8]. FADN is a big database, maintained under auspices of European Commission. Every "commercial" agricultural farm is obliged to send its data to a national branch of FADN (so-called *liaison agent*), which transmit the data to the global FADN database and is responsible for the international comparability of the data (i.e. the methodology of the collecting the data shall be the same in every EU country). A farm is considered commercial if it meets some size criteria.

In total, FADN covers 5 000 000 farms, grouped in 80 000 holdings. These farms maintain about 90 % of the total utilized agricultural area in EU and also produce about 90 % of agricultural output. The farms are categorized

using several criteria, such as the major area of production among others.

The database consists of ~ 1000 economic and other variables on annual basis.

The FADN data are confidential and are not publicly available. However, fortunately, one of the authors has obtained a subset of data for his research under the project of University of Economics, Prague, no. IGA F4/58/2017, with a restricted access and conditions laid on manipulating with the data. For this reason, the data cannot be presented in the paper and we can provide only the results of the computation.

Variables and units.

The set of units with a moderate size was selected. The common property of the units is that they are farms from Czech Republic with livestock production focused on pigs. According to FADN categorization, this is fulfilled by 38 farms.

The criteria were selected to reflect the fact that the farms are focused on *production* – as inputs, four variables are used, corresponding to usual production factors. As the output, the amount of product was used. The variables are summarized in Table 1. Total labour input is measured in *annual work units*. Total farming overheads account for costs linked to production and the workflow of the farm, but *not* linked to a concrete production line. For more detailed description of the variables, see FADN documentation [7].

FADN code	Name	Unit	Type
SE010	Total labour input	AWU	input
SE206	Total output livestock & livestock products	EUR	output
SE336	Total farming overheads	EUR	input
SE365	Total external factors	EUR	input
SE441	Total fixed assets	EUR	input

Table 1 The variables used. A more comprehensive description in text.

3.2 Results

The results of the computational experiments are presented in Table 2. Note that for confidentiality reasons, the units are ordered at random. For every unit, we provide the following: its DEA efficiency E_k, its orthant-specific halfspace depth T_k and its orthant-specific peeling depth P_k. Also, we provide rankings according to these criteria, $r_{E_k}, r_{T_k}, r_{P_k}$. The meaning of rankings is the following. If a unit U_k has a ranking r according to some criterion, it reads that there are $r-1$ units better than U_k.

The results are in compliance with Lemma 1 – if a unit is efficient in the DEA sense, it has depths equal to 1. Also, note that there are only 4 convex hull layers, while the maximal halfspace depth is 11. Some of the units expose the disadvantages of DEA: for example, unit 16 has depths equal to 2, however, its DEA efficiency is only 0.45.

4 Conclusion

We discussed modified halfspace depth as a measure of efficiency of units. We show that a unit is efficient in the DEA sense if and only if it has halfspace depth equal to 1. We also discuss convex hull peeling depth, which is quite similar to DEA; DEA can actually be used as a subprocedure to compute this kind of depth measure.

The modified halfspace depth provides a more robust information about units. In particular, it says how many units have to be removed for a unit to become efficient. On other hand, the modified halfspace depth needs exponential time to be computed.

Acknowledgements

The work of the first and second author was supported by the grant no. 17-13086S of the Czech Science Foundation and by the VŠE IGA project F4/2017/58. The work of the third author was supported by grant no. 16-00408S of Czech Science Foundation. The authors are also grateful to FADN [8], which allowed the second author to access the data necessary for conducting the computational experiments.

References

[1] Avis, D., and Fukuda, K.: Reverse search for enumeration. *Discrete Applied Mathematics* **65** (1996), 21–46.

k	E_k	r_{E_k}	T_k	r_{T_k}	P_k	r_{P_k}
1	1.00	1	1	1	1	1
2	1.00	1	1	1	1	1
3	1.00	1	1	1	1	1
4	0.83	21	2	15	2	15
5	1.00	1	1	1	1	1
6	1.00	1	1	1	1	1
7	0.85	20	3	27	2	15
8	1.00	1	1	1	1	1
9	0.93	15	2	15	2	15
10	0.91	17	6	24	2	15
11	0.73	26	9	33	3	28
12	0.79	24	10	30	2	15
13	0.56	29	11	37	3	28
14	0.48	35	7	31	3	28
15	0.86	19	2	15	2	15
16	0.45	36	2	15	2	15
17	0.58	28	5	27	3	28
18	0.81	23	2	15	2	15
19	0.51	34	2	15	2	15
20	1.00	1	1	1	1	1
21	0.52	32	5	31	3	28
22	0.90	18	3	15	2	15
23	0.52	31	11	27	3	28
24	1.00	1	1	1	1	1
25	0.55	30	7	35	3	28
26	1.00	1	1	1	1	1
27	0.83	22	5	15	2	15
28	1.00	1	1	1	1	1
29	0.41	37	10	37	3	28
30	0.31	38	2	35	4	29
31	1.00	1	1	1	1	1
32	0.51	33	2	34	3	28
33	0.74	25	8	26	3	28
34	0.92	16	2	15	2	15
35	1.00	1	1	1	1	1
36	1.00	1	1	1	1	1
37	0.70	27	4	24	2	15
38	1.00	1	1	1	1	1

Table 2 Results: efficiency (E_k), depths (T_k and P_k), rankings w.r.t. the computed values (denoted by r).

[2] Banker, R. D., Cooper, W. W., Seiford, L. M., Thrall, R. M., and Zhu, J.: Returns to scale in different DEA models. *European Journal of Operational Research* **154** (2004), 345–362.

[3] Barnett, V.: The Ordering of Multivariate Data. *Journal of the Royal Statistical Society. Series A (General)* **139** (1976), 318–355.

[4] Buck, R. C.: Partition of space. *The American Mathematical Monthly* **50** (1943), 541–544.

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

[6] Donoho, D. L., and Gasko, M.: Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness. *The Annals of Statistics* **20** (1992), 1803–1827.

[7] EC: Agriculture - FADN : F. A. D. N. - DEFINITION OF VARIABLES. http://ec.europa.eu/agriculture/rica/definitions_en.cfm, 2018.

[8] EC: Agriculture - FADN : F. A. D. N. - Homepage. http://ec.europa.eu/agriculture/rica/index.cfm, 2018.

[9] Eppstein, D.: Depth and Arrangements. *Invited talk at DIMACS Workshop* (2003).

[10] Frýd, L., and Sokol, O.: Effectivity of the agriculture business in EU: How subsidies affect effectivity? In: *15th International Conference on Data Envelopment Analysis (DEA)*. University of Economics, Prague, 84.

[11] Johnson, D. S., and Preparata, F. P.: The densest hemisphere problem. *Theoretical Computer Science* **6** (1978), 93–107.

[12] Mosler, K.: Depth Statistics. In: *Robustness and Complex Data Structures* (Becker, C., Fried, R., and Kuhnt, S., eds.). Springer Berlin Heidelberg, 2013, 17–34.

[13] Rada, M., and Černý, M.: A New Algorithm for Enumeration of Cells of Hyperplane Arrangements and a Comparison with Avis and Fukuda's Reverse Search. *SIAM Journal on Discrete Mathematics* **32** (2018), 455–473.

[14] Rousseeuw, P. J., and Hubert, M.: Regression Depth. *Journal of the American Statistical Association* **94** (1999), 388–402.

[15] Rousseeuw, P. J., and Hubert, M.: Statistical depth meets computational geometry: A short survey. *arXiv:1508.03828 [stat]* (2015).

[16] Tukey, J. W.: Mathematics and the picturing of data. In: *Proceedings of the International Congress of Mathematicians*, volume 2. 523–531.

[17] Zuo, Y., and Serfling, R.: General notions of statistical depth function. *Annals of statistics* (2000), 461–482.

Condition of Order Preservation in Pairwise Comparisons Matrix With Fuzzy Elements

Jaroslav Ramík[1]

Abstract. In this paper we deal with Preference Order Preservation (POP) condition of pairwise comparisons (PC) matrix with fuzzy elements. Fuzzy elements are appropriate whenever the decision maker (DM) is uncertain about the value of his/her evaluation of the relative importance of elements in question, or, when aggregating crisp pairwise comparisons of a group of decision makers in the group DM problem. We formulate the problem in a general setting investigating pairwise comparisons matrices with fuzzy elements from Abelian linearly ordered group (alo-group). Such an approach enables extensions of traditional multiplicative, additive or fuzzy approaches. Then we propose new order preservation concept based on alpha-cuts. We define the concept of consistency of PC matrix with fuzzy elements (PCF matrices), generalize the concept of the preference order preservation condition (POP condition) to PCF matrices defining weak POP and strong POP condition and derive sufficient conditions for POP condition to be met. Finally, we discuss a numerical example in order to illustrate the proposed concepts and properties.

Keywords: multi-criteria optimization, pairwise comparison matrix, fuzzy elements, strong consistency, alo-group.

JEL classification: C44
AMS classification: 90C15

1 Introduction

The problem we consider here is motivated by a standard *decision making problem (DM problem)*. Let $\mathcal{A} = \{a_1, a_2, ..., a_n\}$ be a finite set of alternatives ($n > 1$). The goal is to *rank the alternatives* from the best to the worst (or, vice versa, which is equivalent), using information given by the decision maker in the form of an $n \times n$ *pairwise comparisons matrix*. The ranking of the alternatives is determined by the *priority vector* of positive numbers $w = (w_1, w_2, ..., w_n)$. This vector, if normalized, is called the *vector of weights*. There exist various methods how to calculate the vector of weights based on the DM problem, particularly, on the pairwise comparisons matrix, see e.g. [3].

Fuzzy sets as the elements of the pairwise comparisons matrix can be applied in the DM problem whenever the decision maker is not sure about the preference degree of his/her evaluations of the pairs in question. Fuzzy elements may be useful also as aggregations of crisp pairwise comparisons of a group of decision makers in the group DM problem. A recent extensive review of various relevant aspects of the pairwise comparison matrices can be found in [3].

The Eigenvalue Method (EVM) introduced by T. Saaty, [7], is, however, criticized by many researchers for some unpleasant properties, see e.g. [1], [8]. The *Preservation of Order Preference condition*, (POP condition for a PC matrix), claims that the ranking result in relation to the given pair of alternatives and priority vector should not break with the expert judgment. Verifying whether for a certain set of pairwise comparisons POP condition is met or not requires rating of alternatives. Therefore, the question arises whether the relationship between POP condition and inconsistency is of a general nature and if so, how does this relationship look like? In the presented work we generalize the concept of POP condition defining α-Weak POP and α-Strong POP condition. Similarly, as in [4], we propose the global error index. Finally, we discuss a numerical example in order to illustrate the proposed concepts and properties.

2 Preliminaries

The fundamental concept of this paper, the *pairwise comparisons matrix with fuzzy elements*, is based on the concept of level sets (or, alpha-cuts) of its elements. Therefore, it will be useful to describe the fuzzy sets as special nested families of subsets of a set, particularly, the families of level sets, for details see [6].

Hence, a *fuzzy subset* of a nonempty set X (or a *fuzzy set* on X) is a family $A = \{A_\alpha\}_{\alpha \in [0;1]}$ of subsets of X satisfying special conditions, see e.g. [5]. The *membership function of* A is the function μ_A from X into the unit

[1]Silesian University in Opava/School of Business Administration in Karvina, Dept. of Informatics and Mathematics, University Sq. 1934/3, 73340 Karvin, Czech Republic, ramik@opf.slu.cz

interval $[0;1]$ defined by $\mu_A(x) = \sup\{\alpha \mid x \in A_\alpha\}$. Given $\alpha \in]0;1]$, the set $[A]_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\}$ is called the α-*cut of fuzzy set* A.

We say that a fuzzy subset A of $\mathbb{R}^* = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ is a *fuzzy interval* whenever its membership function μ_A satisfies the following condition: there exist $a, b, c, d \in \mathbb{R}^*$, $-\infty \leq a \leq b \leq c \leq d \leq +\infty$, such that $\mu_A(t) = 0$ if $t < a$ or $t > d$, μ_A is strictly increasing and continuous on $[a;b]$, $\mu_A(t) = 1$ if $b \leq t \leq c$, and , μ_A is strictly decreasing and continuous on the interval $[c;d]$.

A fuzzy interval A is *bounded* if $[a;d]$ is a bounded interval. For $\alpha = 0$ we define the *zero-cut* of A as $[A]_0 = [a;d]$. A bounded fuzzy interval A is the *triangular fuzzy number* if $b = c$. We denote it by $A = (a, b, d)$. If $b < c$, then A is usually called the *trapezoidal fuzzy number*. We denote it by $A = (a, b, c, d)$. Notice that each crisp number, or, bounded crisp interval, is also a bounded fuzzy interval $A = (a, b, c, d)$. Hence, pairwise comparisons in our approach cover also evaluations of pairs by interval values, (and crisp values).

In order to unify various approaches and prepare a more flexible presentation, we introduce mathematical structures called alo-groups, defined bellow, where only a single binary operation is necessary. An *abelian group*, [2], 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 [2]. 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. 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 $a, b, c \in G : a \leq b$ implies $a \odot c \leq b \odot c$. G 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 $G \times G$.

By the associative property, the operation $\odot$ is extended by induction to n-ary operation. $\mathcal{G}$ 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. $\left(a^{(1/n)}\right)^{(n)} = a$.

In [5], [6], we presented the well known examples of alo-groups, i.e. *Additive alo-group* $\mathcal{R} = (\mathbb{R}, +, \leq)$, *Multiplicative alo-group* $\mathcal{R}_+ = (\mathbb{R}_+, \bullet, \leq)$, *Fuzzy additive alo-group* $\mathcal{R}_a$=$(\mathbb{R}, +_f, \leq)$, and, *Fuzzy multiplicative alo-group* $\mathcal{R}_m$=$(]0;1[, \bullet_f, \leq)$.

3 PCF matrices, consitency, transitivity

Now, we shall investigate an $n \times n$ pairwise comparisons (PC) matrix where each element is a bounded fuzzy interval of the divisible and continuous alo-group $\mathcal{G} = (G, \odot, \leq)$ over an interval G of $\mathbb{R}$, (where $\mathbb{R}$ is the set of real numbers). Moreover, we assume that all diagonal elements of this matrix are crisp, particularly they are equal to the identity element of $\mathcal{G}$. We also assume that this matrix is reciprocal, (which will be defined bellow). We will call this matrix shortly a *PCF matrix*, where the letter *F* denotes the reference to fuzzy elements.

In practice, the type of alo-group will depend on the particular DM problem. Notice that elements of PCF matrices may be crisp and/or fuzzy numbers, and/or fuzzy intervals, and/or fuzzy intervals with bell-shaped membership functions, triangular fuzzy numbers, trapezoidal fuzzy numbers etc. Such fuzzy elements may be either evaluated by individual decision makers, or, they may be made up of crisp pairwise evaluations of decision makers in a group DM problem.

Let $C = \{\tilde{c}_{ij}\}$ be a crisp PCF matrix, simply called PC matrix, where $c_{ij} \in G \subset \mathbb{R}$ for all $i, j \in \{1, 2, ..., n\}$. $C = \{c_{ij}\}$ is said to be $\odot$-*reciprocal*, if the following condition holds: For every $i, j \in \{1, 2, ..., n\}$

$$c_{ij} \odot c_{ji} = e, \text{ or, equivalently, } c_{ji} = c_{ij}^{(-1)}. \tag{1}$$

A crisp PCF matrix $C = \{c_{ij}\}$ is $\odot$-*consistent,* if for all $i, j, k \in \{1, 2, ..., n\}$

$$c_{ik} = c_{ij} \odot c_{jk}, \text{ or, equivalently, } c_{ij} \odot c_{jk} \odot c_{ik}^{(-1)} = e. \tag{2}$$

Remember that an $\odot$-consistent PC matrix $C = \{c_{ij}\}$ is $\odot$-reciprocal, but not vice-versa. The following equivalent condition for consistency of PC matrices is well known, e.g. in AHP, see [7], or [5].

Proposition 1. *A crisp PC matrix* $C = \{c_{ij}\}$ *is* $\odot$*-consistent if and only if there exists a vector* $w = (w_1, w_2, ..., w_n)$, $w_i \in G$, *such that*

$$w_i \div w_j = c_{ij} \text{ for all } i, j \in \{1, 2, ..., n\}. \tag{3}$$

Here, by $\div$ *we define the* inverse operation *satisfying* $w_i \div w_j = w_i \odot w_j^{(-1)}$.

Now, we extend the above stated definition of $\odot$-consistency to non-crisp matrices with fuzzy elements, (see also [6]). Particularly, we introduce a new concept of consistency based on α-cuts: α-$\odot$-consistency. We start, however, with the definition of the α-$\odot$-reciprocity.

Definition 1. Let $\alpha \in [0;1]$, let $C = \{\tilde{c}_{ij}\}$ be a pairwise comparisons matrix with elements being bounded fuzzy intervals. C is said to be α-$\odot$-*reciprocal*, if the following conditions hold:
(i) For each $i \in \{1, 2, ..., n\}$ it holds:

$$c_{ii}^L(\alpha) = c_{ii}^R(\alpha) = e. \tag{4}$$

(ii) For each $i, j \in \{1, 2, ..., n\}$ it holds:

$$c_{ij}^L(\alpha) \odot c_{ji}^R(\alpha) = e. \tag{5}$$

From now on, we shal assume that each PCF matrix $C = \{\tilde{c}_{ij}\}$ is α-$\odot$-reciprocal for all $\alpha \in [0;1]$.

Remark 1. In Definition 1, (i), we assume that all diagonal elements of the PCF matrix are crisp, particularly they are equal to the identity element of $\mathcal{G}$. An interpretation of this assumption is natural: comparing an alternative with itself, the result is *equal*, expressed by the identity element e.
In (ii) we assume that each PCF matrix is α-$\odot$-reciprocal in the usual sense: the reciprocal element of the left side of α-cut is the inverse of the right side of the symmetric α-cut with respect to the operation $\odot$, and vice-versa. Now, we define the concept of consistency of PCF matrices. We start with the definition of α-$\odot$-consistent PCF matrix.

Definition 2. Let $\alpha \in [0;1]$. A PCF matrix $C = \{\tilde{c}_{ij}\}$ is said to be α-$\odot$-*consistent*, if the following condition holds: There exists a crisp matrix $C' = \{c'_{ij}\}$ with $c'_{ik} \in [\tilde{c}_{ik}]_\alpha$, $c'_{ij} \in [\tilde{c}_{ij}]_\alpha$, $c'_{jk} \in [\tilde{c}_{jk}]_\alpha$, such that for each $i, j, k \in \{1, 2, ..., n\}$ it holds

$$c'_{ik} = c'_{ij} \odot c'_{jk}, \text{ or, equivalently, } c'_{ij} \odot c'_{jk} \odot c'^{(-1)}_{ik} = e. \tag{6}$$

The PCF matrix C is said to be $\odot$-*consistent*, if C is α-$\odot$-consistent for all $\alpha \in [0;1]$.
If for some $\alpha \in [0;1]$ the PCF matrix C is not α-$\odot$-consistent, then C is called α-$\odot$-*inconsistent.*

Remark 2. In other words, Definition 2 says that a PCF matrix $C = \{\tilde{c}_{ij}\}$ is α-$\odot$-consistent if there exists a crisp *consistent* matrix $C' = \{c'_{ij}\}$, such that $c'_{ik} \in [\tilde{c}_{ik}]_\alpha$, $c'_{ij} \in [\tilde{c}_{ij}]_\alpha$, $c'_{jk} \in [\tilde{c}_{jk}]_\alpha$.

Remark 3. Let $\alpha, \beta \in [0;1], \alpha \geq \beta$. Evidently, if $C = \{\tilde{c}_{ij}\}$ is α-$\odot$-consistent, then it is β-$\odot$-consistent. The next proposition gives two necessary and sufficient conditions for a PCF matrix to be α-$\odot$-consistent, see also [6]. Condition (ii) is an extension of Proposition 1.

Proposition 2. *Let $C = \{\tilde{c}_{ij}\}$ be a PCF matrix, $\alpha \in [0;1]$, $[\tilde{c}_{ij}]_\alpha = [c_{ij}^L(\alpha), c_{ij}^R(\alpha)]$ be an α-cut, $i, j \in \{1, 2, ..., n\}$. The following three conditions are equivalent.*
(i) $C = \{\tilde{c}_{ij}\}$ is α-$\odot$-consistent.
(ii) There exists a vector $w = (w_1, w_2, ..., w_n)$ with $w_i \in G, i \in \{1, 2, ..., n\}$, such that for each $i, k \in \{1, 2, ..., n\}$, it holds:

$$c_{ik}^L(\alpha) \leq w_i \div w_k \leq c_{ik}^R(\alpha). \tag{7}$$

(iii) For each $i, j, k \in \{1, 2, ..., n\}$, it holds:

$$c_{ik}^L(\alpha) \leq c_{ij}^R(\alpha) \odot c_{jk}^R(\alpha), \tag{8}$$

$$c_{ik}^R(\alpha) \geq c_{ij}^L(\alpha) \odot c_{jk}^L(\alpha). \tag{9}$$

Remark 4. Property (iii) in Proposition 2 is useful for checking α-$\odot$-consistency of PCF matrices. For a given PCF matrix $C = \{\tilde{c}_{ij}\}$ it can be easily calculated whether inequalities (8) and (9) are satisfied or not.

4 Preservation of Order Preference (POP) condition for PCF matrix

The goal of the pairwise comparisons method is ranking the alternatives. For this purpose, we set up a function that assigns positive values to the alternatives. After that, ranking (and/or ordering) the alternatives is a simple task.

Definition 3. Let $\mathcal{A}$ be a finite set of alternatives $\mathcal{A} = \{a_1, a_2, ..., a_n\}$. The *ranking function of* $\mathcal{A}$ (the *ranking of* $\mathcal{A}$) is a function $w : \mathcal{A} \to \mathbb{R}_+$ that assigns to every alternative from $\mathcal{A}$ a positive value from $\mathbb{R}_+$. We denote it by vector $w = (w(a_1), ..., w(a_n))$.

In other words, $w(a)$ represents the ranking value for $a \in \mathcal{A}$. The w function is usually denoted in the form of a vector of weights $w = (w_1, ..., w_n)$ and it is called the *priority vector*.

Let $C = \{c_{ij}\}$ be a crisp PCF matrix on alo-group $\mathcal{G} = (G, \odot, \leq)$, i.e. $c_{ij} \in G$ for all $i, j \in \{1, 2, ..., n\}$, called simply PC matrix, and let $a_i, a_j \in \mathcal{A}$. Let us define the relations of dominance $\succ$ and $\simeq$ on the set $\mathcal{A}$ of alternatives by:

$$a_i \succ a_j \text{ if } c_{ij} > e, a_i \simeq a_j \text{ if } c_{ij} = e. \tag{10}$$

If $a_i \succ a_j$, we say that a_i *strictly dominates* a_j , whereas if $a_i \simeq a_j$ we say that a_i and a_j are *indifferent.* We denote $a_i \succeq a_j$ if ($a_i \succ a_j$ or $a_i \simeq a_j$). If $a_i \succeq a_j$, we say that a_i *dominates* a_j.

Definition 4. A (crisp) PC matrix $C = \{c_{ij}\}$ is $\odot$-*transitive*, if the following conditions (11) and (12) hold:

$$\text{if } c_{ij} > e \text{ and } c_{jk} > e, \text{ then } c_{ik} > e, \tag{11}$$

$$\text{if } c_{ij} = e \text{ and } c_{jk} = e, \text{ then } c_{ik} = e. \tag{12}$$

Proposition 3. *Let $C = \{c_{ij}\}$ be $\odot$-transitive, then there exists a priority vector $w = (w(a_1), ..., w(a_n))$, with $w(a_i) \in G$ for all $i \in \{1, 2, ..., n\}$, such that*

$$a_i \succeq a_j \text{ if and only if } w(a_i) \geq w(a_j). \tag{13}$$

In [1] *Bana e Costa* and *Vansnick* formulate two conditions of order preservations for PC matrix, which is not necessarily consistent, or, transitive.

The *Preservation of Order Preference condition*, (POP condition for a PC matrix $C = \{c_{ij}\}$ and priority vector w), claims that the ranking result in relation to the given pair of alternatives a_i, a_j and priority vector w should not break with the expert judgment, i.e.:
If for a pair of alternatives $a_i, a_j \in \mathcal{A}$ it holds $a_i \succ a_j$, or, equivalently, $c_{ij} > e$, then

$$w(a_i) > w(a_j), \text{ or, equivalently, } w(a_i) \div w(a_j) > e. \tag{14}$$

The POP condition (14) is, in fact, weaker than (13), in Proposition 3, which holds, however, for transitive matrices and corresponding priority vectors. Now, we are going to define the POP condition for a PCF matrix $C = \{\tilde{c}_{ij}\}$.

Definition 5. Let $a_i, a_j \in \mathcal{A}$, $C = \{\tilde{c}_{ij}\}$ be a PCF matrix on alo-group $\mathcal{G} = (G, \odot, \leq)$, $\alpha \in [0; 1]$.
(i) a_i *α-weakly dominates* a_j, if there exists $c_{ij} \in [\tilde{c}_{ij}]_\alpha$, such that it holds $c_{ij} > e$.
Moreover, a_i *weakly dominates* a_j, if a_i α-weakly dominates a_j, for all $\alpha \in [0; 1]$.
(ii) a_i *α-strongly dominates* a_j, if for each $c_{ij} \in [\tilde{c}_{ij}]_\alpha$, it holds $c_{ij} > e$.
Moreover, a_i *strongly dominates* a_j, if a_i α-strongly dominates a_j, for all $\alpha \in [0; 1]$.

Proposition 4. *Let $a_i, a_j \in \mathcal{A}$, $C = \{\tilde{c}_{ij}\}$ be a PCF matrix on alo-group $\mathcal{G}$, $\alpha \in [0; 1]$, let $[\tilde{c}_{ij}]_\alpha = [c_{ij}^L(\alpha), c_{ij}^R(\alpha)]$ be an α-cut.*
(i) a_i *α-weakly dominates* a_j, *if and only if* $c_{ij}^R(\alpha) > e$.
(ii) a_i *weakly dominates* a_j, *if and only if* $c_{ij}^R(1) > e$.
(iii) a_i *α-strongly dominates* a_j, *if and only if* $c_{ij}^L(\alpha) > e$.
(iv) a_i *strongly dominates* a_j, *if and only if* $c_{ij}^L(0) > e$,

Definition 6. Let $a_i, a_j \in \mathcal{A}$, $C = \{\tilde{c}_{ij}\}$ be a PCF matrix on alo-group $\mathcal{G} = (G, \odot, \leq)$, $w = (w_1, w_2, ..., w_n)$, $w_i \in G$, be a priority vector, $\alpha \in [0; 1]$.
(i) We say that the *α-weak preservation of order preference condition (α-WPOP condition)* is satisfied:
if a_i α-weakly dominates a_j, then $w_i > w_j$.
(ii) We say that the *weak preservation of order preference condition (WPOP condition)* is satisfied:
if a_i weakly dominates a_j, then $w_i > w_j$.
(iii) We say that the *α-strong preservation of order preference condition (α-SPOP condition)* is satisfied:
if a_i α-strongly dominates a_j, then $w_i > w_j$.
(iv) We say that the *strong preservation of order preference condition (SPOP condition)* is satisfied:
if a_i strongly dominates a_j, then $w_i > w_j$.

Remark 5. Notice, that the concepts of satisfaction of α-weak or α-strong preservation of order preference condition (i.e. α-WPOP condition, resp. α-SPOP condition) is introduced for a given PCF matrix C and given priority vector w. If $C = \{\tilde{c}_{ij}\}$ is a crisp PCF matrix, then α-WPOP, WPOP, α-SPOP, and SPOP conditions coincide with the usual POP condition (14).
By Definition 6, WPOP condition implies α-WPOP condition for any $\alpha \in [0; 1]$. Analogically, SPOP condition implies α-SPOP condition for any $\alpha \in [0; 1]$. It is evident, that α-SPOP condition implies α-WPOP condition, where $\alpha \in [0; 1]$, but not vice versa. Analogically, SPOP condition implies WPOP condition, but not vice versa.

Definition 7. Let $C = \{\tilde{c}_{ij}\}$ be a PCF matrix on alo-group $\mathcal{G} = (G, \odot, \leq)$. For each pair $i, j \in \{1, 2, \ldots, n\}$, and a priority vector $w = (w_1, w_2, ..., w_n)$, $w_i \in G$, and $\alpha \in [0; 1]$. Let us denote:

$$\epsilon^L(i, j, w, \alpha) = c_{ij}^L(\alpha) \odot w_j \div w_i, \quad \epsilon^R(i, j, w, \alpha) = c_{ij}^R(\alpha) \odot w_j \div w_i. \tag{15}$$

Further, let

$$\varepsilon^L(i, j, w, \alpha) = \max\{\epsilon^L(i, j, w, \alpha), (\epsilon^L(i, j, w, \alpha))^{(-1)}\}, \tag{16}$$

$$\varepsilon^R(i, j, w, \alpha) = \max\{\epsilon^R(i, j, w, \alpha), (\epsilon^R(i, j, w, \alpha))^{(-1)}\}. \tag{17}$$

Moreover, define the $n \times n$ matrices of *local error indexes* $\varepsilon^L(w, \alpha)$ and $\varepsilon^R(w, \alpha)$ as

$$\varepsilon^L(w, \alpha) = \{\varepsilon^L(i, j, w, \alpha)\}, \quad \varepsilon^R(w, \alpha) = \{\varepsilon^R(i, j, w, \alpha)\}. \tag{18}$$

By reciprocity of C we obtain

$$\varepsilon^L(w,\alpha) = \varepsilon^R(w,\alpha)^T. \tag{19}$$

Definition 8. The *global error index* $\mathcal{E}(C,w,\alpha)$, for PCF matrix $C=\{\tilde{c}_{ij}\}$ and a priority vector $w=(w_1,\ldots,w_n)$, is defined as the maximal element of matrix of local errors $\varepsilon^L(w,\alpha)$, i.e.

$$\mathcal{E}(C,w,\alpha) = \max_{i,j\in\{1,\ldots,n\}} \varepsilon^L(i,j,w,\alpha), \tag{20}$$

Now, we present the main results of this paper, the proofs are omitted.

Proposition 5. *Let $C=\{\tilde{c}_{ij}\}$ be a PCF matrix, $i,j\in\{1,2,\ldots,n\}$, $\alpha\in[0;1]$.*
(i) If C is α-$\odot$-consistent, then there exists a priority vector $w(\alpha)=(w_1(\alpha),w_2(\alpha),\ldots,w_n(\alpha))$, such that α-SPOP condition is satisfied, i.e. $c_{ij}^L(\alpha)>e$ implies $w_i>w_j$.
(ii) If C is $\odot$-consistent, then there exists a priority vector $w=(w_1,w_2,\ldots,w_n$, such that SPOP condition is satisfied, i.e. $c_{ij}^L(0)>e$ implies $w_i>w_j$.

The following proposition gives a sufficient condition for satisfying the SPOP and WPOP conditions with respect to the value of global error index (20).

Proposition 6. *Let $C=\{\tilde{c}_{ij}\}$ be a PCF matrix, $w=(w_1,w_2,\ldots,w_n)$ be a priority vector, $i,j\in\{1,2,\ldots,n\}$.*
(i) *If $c_{ij}^L(\alpha)>\mathcal{E}(C,w,\alpha)$, then α-SPOP condition is satisfied, i.e. $c_{ij}^L(\alpha)>e$ implies $w_i>w_j$.*
(ii) *If $c_{ij}^R(\alpha)>\mathcal{E}(C,w,\alpha)$, then α-WPOP condition is satisfied, i.e. $c_{ij}^R(\alpha)>e$ implies $w_i>w_j$.*

5 Priority vectors, measuring inconsistency

In the previous propositions, there is no assumption concerning the method for generating priority vector $w=(w(a_1),\ldots,w(a_n))$. Now, we propose a general method for calculating the priority vector of PCF matrix for rating the alternatives. This method is based on Proposition 2, (ii), particularly on the optimal solution of the following optimization problem:

$$(P) \qquad \alpha \longrightarrow \max; \tag{21}$$

subject to

$$c_{ij}^L(\alpha) \le w_i \div w_j \le c_{ij}^R(\alpha) \text{ for all } i,j\in\{1,2,...,n\}, \tag{22}$$

$$\bigodot_{k=1}^{n} w_k = e, \tag{23}$$

$$0\le\alpha\le 1, w_k\in G, \text{ for all } k\in\{1,2,...,n\}. \tag{24}$$

If optimization problem (*P*) has a feasible solution, i.e. system of constraints (22) - (24) has a solution, then (*P*) has also an optimal solution. Let α^* and $w^*=(w_1^*,...,w_n^*)$ be an optimal solution of problem (*P*). Then α^* is called the $\odot$-*consistency grade of PCF matrix* C, denoted by $g_\odot(C)$, i.e. $g_\odot(C)=\alpha^*$. Here, $w^*=(w_1^*,...,w_n^*)$ is the $\odot$-priority vector of PCF matrix C.

If optimization problem (*P*) has no feasible solution, then we define $g_\odot(C)=0$. In that case, the priority vector is not defined. Then either a procedure described in [5], or, [6] can be applied, or, the DM is asked to reconsider the problem and newly evaluate some elements of PCF matrix.

6 Example

Consider the multiplicative alo-group $\mathcal{R}_+=(\mathbb{R}_+,\bullet,\le)$ with $\odot=\bullet$, where $\bullet$ is the usual operation of multiplication. Let PCF matrices $C=\{\tilde{c}_{ij}\}$ be given as follows:

$$C=\begin{bmatrix} (1,1,1,1) & (\frac{2}{3},1,2,3) & (6,7,8,9) \\ (\frac{1}{3},\frac{1}{2},1,\frac{3}{2},) & (1,1,1,1) & (2,3,9,12) \\ (\frac{1}{9},\frac{1}{8},\frac{1}{7},\frac{1}{6},) & (\frac{1}{12},\frac{1}{9},\frac{1}{3},\frac{1}{2},) & (1,1,1,1) \end{bmatrix},$$

or, equivalently, using α-cuts, for $\alpha\in[0;1]$ we obtain

$$C=\begin{bmatrix} [1;1] & [\frac{2+\alpha}{3};3-\alpha] & [6+\alpha;9-\alpha] \\ [\frac{1}{3-\alpha};\frac{3}{2+\alpha}] & [1;1] & [2+\alpha;12-3\alpha] \\ [\frac{1}{9-\alpha};\frac{1}{6+\alpha}] & [\frac{1}{12-3\alpha};\frac{1}{2+\alpha}] & [1;1] \end{bmatrix}.$$

Here, C is a 3×3 PCF matrix, particularly a PCF matrix with trapezoidal fuzzy elements, and the corresponding piecewise linear membership functions.
Solving problem (*P*), it can be easily shown that C is α-•-consistent for $0 \leq \alpha \leq \alpha^* = 0,828$.
The priority vector is calculated as $w^* = (0,575; 0,314; 0,111)$.
First, we calculate the 3×3 matrices of local error indexes by (15) $\varepsilon^L(w^*, \alpha^*)$ and $\varepsilon^R(w^*, \alpha^*)$ as

$$\{\epsilon^L(i,j,w^*,\alpha^*)\} = \begin{bmatrix} 1 & \frac{2+\alpha^*}{3}\cdot\frac{w_2^*}{w_1^*} & (6+\alpha^*).\frac{w_3^*}{w_1^*} \\ \frac{1}{3-\alpha^*}\cdot\frac{w_1^*}{w_2^*} & 1 & (2+\alpha^*).\frac{w_3^*}{w_2^*} \\ \frac{1}{9-\alpha^*}\cdot\frac{w_1^*}{w_3^*} & \frac{1}{12-3\alpha^*}\cdot\frac{w_2^*}{w_3^*} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.515 & 1.318 \\ 0.843 & 1 & 1.000 \\ 0.634 & 3.949 & 1 \end{bmatrix},$$

and

$$\{\epsilon^R(i,j,w^*,\alpha^*)\} = \begin{bmatrix} 1 & (3-\alpha^*).\frac{w_2^*}{w_1^*} & (9-\alpha^*).\frac{w_3^*}{w_1^*} \\ \frac{3}{2+\alpha^*}\cdot\frac{w_1^*}{w_2^*} & 1 & (12-3\alpha^*).\frac{w_3^*}{w_2^*} \\ \frac{1}{6+\alpha^*}\cdot\frac{w_1^*}{w_3^*} & \frac{1}{2+\alpha^*}\cdot\frac{w_2^*}{w_3^*} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1.186 & 1.577 \\ 1.942 & 1 & 3.364 \\ 0.7598 & 1.000 & 1 \end{bmatrix}.$$

By (18) we obtain the left and right matrices of local error indexes, $\varepsilon^L(w^*, \alpha^*)$ and $\varepsilon^R(w^*, \alpha^*)$, as

$$\varepsilon^L(w^*,\alpha^*) = \{\varepsilon^L(i,j,w^*,\alpha^*)\} = \begin{bmatrix} 1 & 1.942 & 1.318 \\ 1.186 & 1 & 1.000 \\ 1.577 & 3.949 & 1 \end{bmatrix} = \{\varepsilon^R(i,j,w^*,\alpha^*)\}^T.$$

Notice, that $\varepsilon^L(w^*, \alpha^*) = \varepsilon^R(w^*, \alpha^*)^T$. The global error index $\mathcal{E}^L(C, w^*, \alpha^*)$, for the PCF matrix C, priority vector w^* and α^* is calculated by (20) as $\mathcal{E}(C, w^*, \alpha^*) = 3.363$.

7 Conclusion

This paper deals with pairwise comparisons matrix with fuzzy elements which consists of pairwise comparisons of a finite set of alternatives. We generalized the concept of the preference order preservation condition (POP condition) to PCF matrices defining α-Weak POP and α-Strong POP and derived several sufficient conditions for POP condition to be met. Finally, we discuss a numerical example in order to illustrate the proposed concepts and properties.

Acknowledgements

This research has been supported by GACR project No. 18-01246S.

References

[1] Bana e Costa, C.A., Vasnick, J. C. (2008). A critical analysis of the eigenvalue method used to derive priorities in the AHP. *European Journal of Operational Research*, 187, 3, 1422–1428.

[2] Bourbaki, N. (1998). *Algebra II.* Springer Verlag, Heidelberg-New York-Berlin, 1990.

[3] Kou, G., Ergu, D., Lin, C. S. & Chen, Y. (2016). Pairwise comparison matrix in multiple criteria decision making. *Technological and Economic Development of Economy*, 22, 5, 738–765.

[4] Kułakowski, K. (2015). A heuristic rating estimation algorithm for the pairwise comparisons method. *Central European Journal of Operations Research*, 23, 1, 187–203.

[5] Ramik, J. (2015). Pairwise comparison matrix with fuzzy elements on alo-group. *Information Sciences*, 297, 236–253.

[6] Ramik, J. (2016). Strong Consistent Pairwise Comparisons Matrix with Fuzzy Elements and Its Use in Multi-Criteria Decision Making. In *Proceedings of 34th International Conference Mathematical Methods in Economics 2016* (pp. 723–728). Liberec: Union of Czech Mathematicians and Physicists.

[7] Saaty, T. L., (1977). A Scaling Method for Priorities in Hierarchical Structure. *Journal of Mathematical Psychology*, 15, 3, 234–281.

[8] Whitaker, R. (2007). Criticisms of the Analytic Hierarchy Process: Why they often make no sense. *Mathematical and Computer Modeling*, 46, 7/8, 948–961.

Taxonomic Measure of Development with Entropy Weights in Assessment of Entrepreneurial Conditions in Poland

Elżbieta Rogalska[1]

Abstract. Entrepreneurial conditions are an important factor supporting regional growth and regional convergence process. Good conditions for entrepreneurs are an effective stabiliser of regional economy and contribute to its sustainable socio-economic development. Therefore, the research regarding entrepreneurial conditions is an important scientific problem. In the current article the pointed phenomenon was analysed for Poland at regional level (NUTS 2). The entrepreneurial conditions were assessed with application of five diagnostic variables provided by Statics Poland (Local data bank). Therefore, the entrepreneurial conditions were treated as a multiple-criteria problem. In the multiple-criteria research a dilemma relating to applying appropriate weights for variables is often the source of controversy. As a result, in current study in order to propose objective rating of NUTS 2 regions a taxonomic measure of development with entropy weights was used. The research was done for the years 2011 and 2017. The current research enabled to verify the differences in importance of the diagnostic variables. Additionally, it showed stable in time disparities between regions in Poland, which should be taken in into consideration in the process of forming regional development policy.

Keywords: entrepreneurship, Poland, NUTS 2 regions, entropy weights, taxonomic measure of development

JEL classification: C38, R11

AMS classification: 91C20

1 Introduction

Entrepreneurial conditions are an important factor influencing regional growth. Favourable conditions for entrepreneurs can help to stabilise regional economy during the crisis and contribute to its sustainable socio-economic development and convergence process, which is especially important for Central and Eastern European economies [19; 40].

Therefore, the main aim of the article is to assess entrepreneurial conditions in Poland at NUTS 2 level in the years 2011 and 2017. The current research is a continuation of previous studies of the author, where in [34] the entrepreneurship conditions at NUTS 3 level were analysed with application of zero-unitarization method, and in [35] this phenomenon was quantified with application of taxonomic measure of development based on median vector Weber. In the case of both mentioned papers in the multiple-criteria analysis no weights for the diagnostic variables were applied. However, there are many arguments in favour of the view that in the case of entrepreneurial conditions not all factors forming the phenomenon have the same importance [32; 45]. Therefore, some proposals suggesting approaches for assessing weights for the research concerning entrepreneurial conditions are needed. As a result, in current paper taxonomic measure of development based on Hellwig's method with entropy weights was applied. Thus, the main contribution of the article relates to fact that it enabled to verify the differences in importance of the diagnostic variables, which can be used for assessing entrepreneurial conditions in Poland at regional level.

2 Taxonomic measure of development with entropy weights

The article concentrates on the problem of quantifying entrepreneurship conditions in Polish NUTS 2 regions. The literature review concerning factors influencing the entrepreneurship conditions enables to point at vide vector of determinants that can influence entrepreneurship, which can relate to obstacles and support for SMEs [12; 13; 16; 10], effectiveness of financial markets influencing availability of financing [18; 6; 24; 25; 26]; situation on labour market, which creates incentives for being entrepreneur [14; 29; 33; 36]; attitudes and quality of human capital in a country [42; 2; 37; 22], innovative potential and availability of digital infrastructure [9; 5; 7], effectiveness of institutions both at regional and national level [11; 20; 21; 4].

The provided literature review indicates that the entrepreneurship and entrepreneurial conditions should be considered as a multiple-criteria phenomenon [23; 17; 44; 1; 41]. In order to measure this type of phenomena, a

[1] Elżbieta Rogalska, PhD, University of Wamia and Mazury in Olsztyn, Department of Microeconomics, ul. Plac Cieszyński 1, 10-719 Olsztyn, Poland, e-mail: ela.n.rogalska@gmail.com

taxonomic measure of development (TMD) proposed by Hellwig [15] can be applied. TMD is a tool, which belongs to the group of linear ordering methods [8]. The values of TMD measure in current research are assesed by Hellwig's procedure, which is modified with application of entropy weights [3; 31; 43; 39; 27]. The modification enabled to assess the differences in importance of diagnostic variables, which can be used to describe entrepreneurship conditions at regional level. The applied pocedure of entropy weights is based on the definition of Shannon entropy [38; 3; 31].

The Hellwig's procedure with entropy weights can be given with the following steps: [3; 31]:

1. Adoption of a set of examined objects O_i ($i=1,2,...,m$) and a set of diagnostic variables X_j ($j=1,2,..,n$; $t=1,2,...,T$) for the analysed phenomenon.
2. Transferring of all the diagnostic variables into stimulants [15].
3. Assessing of a set of entropy weights[2] w_j for variable X_j based on the entropy value e_j [43], which can be given with the following equations:

$$p_{ijt} = \frac{x_{ijt}}{\sum_{i=1}^{m} x_{ijt}}, \quad e_{jt} = -\frac{1}{\ln m}\sum_{i=1}^{m} p_{ijt} \ln p_{ijt}, \quad w_{jt} = \frac{1-e_{jt}}{n-\sum_{j=1}^{n} e_{jt}} \tag{1}$$

4. Standardization of diagnostic variables X_j, which can be done with application of classic standardisation based on the mean value and standard deviation. As a result, the set of variables Z_j is obtained. Then a set of variables with weights Y_j is assessed with the equation:

$$y_{ijt} = z_{ijt} w_{jt} \tag{2}$$

5. Assessment for every object O_i a distance to the pattern of development for a given time t with the equation:

$$d_{i0t} = \sqrt{\sum_{j=1}^{p} (y_{ijt} - y_{0jt})^2} \tag{3}$$

where the pattern of development for the variables is assesed with the formlula:

$$y_{0jt} = \max_{it} y_{ijt} . \tag{4}$$

6. Assessment for every object O_i value of TMD_{it} with application of the equation:

$$d_{it} = 1 - \frac{d_{i0t}}{d_{0t}} \tag{5}$$

where $d_{0t} = \bar{d}_{0t} + 2s_{dt}$, and $\bar{d}_{0t}$, s_{dt} are given with formula:

$$\bar{d}_{ot} = \frac{1}{n}\sum_{i=1}^{n} d_{i0t}, \quad s_{ot} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (d_{i0t} - \bar{d}_{ot})^2} \tag{6}$$

3 Assessment of Weights

In order to calculate TMD with entropy weights for entrepreneurship conditions, five diagnostic variables were determined, which are presented in Table 1. The selection of the variables was based on the previous literature review [34; 35] and the literature presented in the previous section. All the variables should be considered as stimulants. The main restriction for the selection of the variables was the availability of data for Poland at NUTS 2 level. The data was obtained form Statics Poland (Local data bank).

Next, on the basis of the variables values separately for the year 2011 and 2017 weights based on entropy value were determined (Table 2). For both periods, the largest weights were assigned to variable x_5 (Gross value of fixed assets in enterprises per capita) and variable x_4 (Capital expenditures in enterprises per capita). It is

[2] The sum of entropy weights w_j is equal to 1.

interesting that in the year 2011 the variable x_5 had the highest influence on entrepreneurship conditions, and in 2017 the variable x_4, which would can indicate a change in the economic and institutional situation in Poland.

Variable	Character	Description
x_1	Stimulant	Number of entities included in the REGON registration per 10 thousand inhabitants
x_2	Stimulant	Natural persons conducting economic activity per 1 thousand inhabitants
x_3	Stimulant	Share of new registered companies form creative industry in the total number of new registered commercial law companies
x_4	Stimulant	Capital expenditures in enterprises per capita
x_5	Stimulant	Gross value of fixed assets in enterprises per capita

Table 1 Diagnostic variables for entrepreneurial conditions in Poland at regional level

2011 year				
x_1	x_2	x_3	x_4	x_5
0,114	0,091	0,095	0,276	0,424
2017 year				
x_1	x_2	x_3	x_4	x_5
0,093	0,065	0,059	0,494	0,289

Table 2 Entropy weights for the year 2011 and 2017

4 Empirical results

In accordance with the adopted objective of the article, an analysis of entrepreneurial conditions in Poland at NUTS 2 level was carried out for the years 2011 and 2017. In order to quantify the phenomenon TMD was used, where weights for particular variables were determined on the basis of entropy value. The values of the TMD were determined based on equations 3-6 for the Hellwig's procedure, where the standardization of variables was based on the arithmetic mean and standard deviation. For both years 2011 and 2017 a constant values of pattern of development was applied, which was a condition for comparability of the results. The obtained TMD allowed to rank the NUTS 2 regions in terms of the entrepreneurship conditions. Then, on the basis of the natural breaks method, the regions were assigned to three classes (Table 2).

2011 year				2017 year		
NUTS 2 Region	**TMD**	**Rank**	**Class**	**TMD**	**Rank**	**Class**
MAZOWIECKIE	0.881	1	1	0.904	1	1
DOLNOŚLĄSKIE	0.598	2	2	0.796	2	1
OPOLSKIE	0.316	9	2	0.770	3	1
WIELKOPOLSKIE	0.428	4	2	0.650	4	2
ŚLĄSKIE	0.492	3	2	0.648	5	2
POMORSKIE	0.427	5	2	0.555	6	2
ŁÓDZKIE	0.381	6	2	0.529	7	2
LUBUSKIE	0.328	8	2	0.489	8	2
MAŁOPOLSKIE	0.296	10	3	0.451	9	2
KUJAWSKO-POMORSKIE	0.340	7	2	0.321	10	3
PODKARPACKIE	0.247	13	3	0.315	11	3
ZACHODNIOPOMORSKIE	0.274	11	3	0.313	12	3
PODLASKIE	0.241	14	3	0.239	13	3
WARMIŃSKO-MAZURSKIE	0.154	15	3	0.226	14	3
LUBELSKIE	0.150	16	3	0.177	15	3
ŚWIĘTOKRZYSKIE	0.256	12	3	0.152	16	3

Table 2 TMD values for entrepreneurial conditions in Poland at regional level

Both in 2011 and 2017 to the first class with the best entrepreneurship conditions Mazowieckie province was assigned. In this region the capital city of Poland is located, which naturally makes it economically the most attractive NUTS. It should be emphasized that the situation regarding entrepreneurship conditions in Poland improved in the analysed period, and in 2017 Dolnośląskie and Opolskie were classified in the class 1. In the class 2 with an average entrepreneurship conditions, in 2011 the following NUTS were found: Dolnośląskie, Opolskie, Wielkopolskie, Śląskie, Pomorskie, Łódzkie, Lubuskie and Kujawsko-pomorskie. NUTS 2 with the worst entrepreneurship conditions have been assigned to class 3. In the class 3 in 2011, the following NUTS were found Podlaskie, Warmińsko-mazurskie, Lubelskie and Świętokrzyskie. Most of these NUTS are located on the Eastern part of Poland, which confirms the traditional economic disparities in the country [28; 30; 31]. In 2017, there were slight changes in class 3, Małopolskie was rated in the class 2, while Kujawsko-pomorskie was degraded to grade 3.

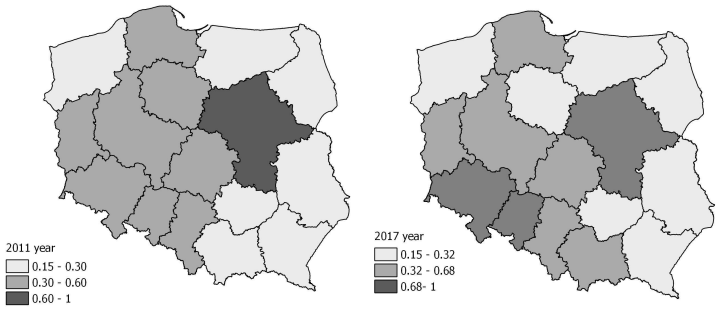


Figure 1 Entrepreneurial conditions in Poland at regional level in the years 2011 and 2017

5 Conclusions

The objective of the research was to assess entrepreneurial conditions in Poland at NUTS 2 level in the years 2011 and 2017. To reach that aim taxonomic measure of development method based on the procedure proposed by Hellwig with entropy weights was applied. The main empirical contribution of the article relates to the verification of the differences in importance of the diagnostic variables, which can be applied for assessing entrepreneurial conditions in Poland at regional level.

From the policy guidelines perspective the article confirmed relatively stable regional disparities between Western and Eastern part of Poland in regard to entrepreneurial conditions. It can indicate that the funds devoted to regional policy with the objective of reducing the disparities between NUTS 2 regions in Poland do not bring the results, which are often expected. As a result, in the future the problem of structural disparities in entrepreneurial conditions can become a significant growth burden for Polish economy, therefore it should be the issue of special attention for Polish policy decision makers both at national and regional level.

References

[1] Ahmed, R.R., Vveinhardt, J. & Streimikiene, D. (2017). Interactive digital media and impact of customer attitude and technology on brand awareness: evidence from the South Asian countries. *Journal of Business Economics and Management*, 18(6), 1115-1134.

[2] Balcerzak, A. P. (2016). Multiple-criteria Evaluation of Quality of Human Capital in the European Union Countries. *Economics & Sociology*, 9(2), 11-27.

[3] Balcerzak, A. P. (2017). Digital Economy in Czech Republic, Slovakia and Hungary. Measurement with TOPSIS Based on Entropy Measure for Objective Weighting. In T. Loster & T. Pavelka (Eds.). *The 11th International Days of Statistics and Economics. Conference Proceedings. September 8-10, 2016*. Prague: Libuse Macakova, Melandrium, 49-57.

[4] Balcerzak, A. P. & Pietrzak, M.B. (2016). 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), 66-81.

[5] Balcerzak, A. P. & Pietrzak, M. B. (2017). Digital Economy in Visegrad Coutnries. Multiple-criteria decision analysis at Regional Level in the Years 2012 and 2015. *Journal of Competitiveness*, 9(2), 5-18.

[6] Balcerzak, A. P., Kliestik, T., Streimikiene, D. & Smrčka L. (2017). Non-parametric approach to measuring the efficiency of banking sectors in European Union Countries. *Acta Polytechnica Hungarica*, 14(7), 51-70.

[7] Begonja, M., Čićek, F., Balboni, B. & Gerbin, A. (2016). Innovation and business performance determinants of SMEs in the Adriatic region that introduced social innovation. *Economic Research-Ekonomska Istraživanja*, 29(1), 1136-1149.

[8] Cheba, K. & Szopik-Depczyńska, K. (2017). Multidimensional comparative analysis of the competitive capacity of the European Union countries and geographical regions. *Oeconomia Copernicana*, *8*(4), 487-504.

[9] Czarniewski, S. (2016). Small and medium-sized enterprises in the context of innovation and entrepreneurship in the economy. *Polish Journal of Management Studies*, 13(1), 30-39.

[10] Dobeš K., Kot S., Kramoliš J. & Sopková G. (2017). The Perception of Governmental Support in The Context of Competitiveness of SMEs in the Czech Republic. *Journal of Competitiveness*, 9(3), 34-50.

[11] Duh, M., Belak, J. & Milfelner, B. (2016). The importance of culture for enterprise dynamics: the role of type and strength of culture. *Economic Research-Ekonomska Istraživanja*, 29(1), 263-285.

[12] Erkut, B. (2016). Entrepreneurship and Economic Freedom: Do Objective and Subjective Data Reflect the Same Tendencies? *Entrepreneurial Business and Economics Review*, 4(3), 11-26.

[13] Fabuš, M. (2017). Current development of business environment in Slovakia and Czech Republic. *Entrepreneurship and Sustainability Issues*, 5(1), 127-137.

[14] Hadaś-Dyduch, M., Balcerzak, A. P. & Pietrzak, M.B. (2016). Wavelet Analysis of Unemployment Rate in Visegrad Countries. In T. Kliestik (Ed.). *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.).* Zilina: University of Zilina, 595-602.

[15] Hellwig, Z. (1972). Procedure of Evaluating High-Level Manpower Data and Typology of Countries by Means of the Taxonomic Method. In Z. Gostowski (ed.), *Towards a System of Human Resources Indicators for Less Developed Countries, Papers Prepared for a UNESCO Research Project*. Wrocław: Ossolineum-The Polish Academy of Sciences, 115-134.

[16] Ivanová, E. (2017). Barriers to the development of SMEs in the Slovak Republic. *Oeconomia Copernicana*, *8*(2), 255-272.

[17] Keshavarz Ghorabaee, M., Zavadskas, E.K., Amiri, M. & Antucheviciene, J. (2016). Evaluation by an Area-based Method of Ranking Interval Type-2 Fuzzy Sets (EAMRIT-2F) for Multi-criteria Group Decision-making. *Transformations in Business & Economics*, 15(3)(39), 76-95.

[18] Kljucnikov, A., & Belas, J. (2016). Approaches of Czech entrepreneurs to debt financing and management of credit risk. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *11*(2), 343-365.

[19] Kuc M. (2017). Is the Regional Divergence a Price for the International Convergence? The Case of the Visegrad Group. *Journal of Competitiveness*, 9(4), 50-65.

[20] Lizińska, W., Marks-Bielska, R., Babuchowska, K. & Wojarska, M. (2016). Institutional efficiency of local government in issuing administrative decisions, exemplified by the performance of the local government appeal board in Olsztyn. *Oeconomia Copernicana*, *7*(2), 285-296.

[21] Lizińska, W., Marks-Bielska, R., Babuchowska, K. & Wojarska, M. (2017). Factors contributing to the institutional efficiency of local governments in the administrative area. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *12*(2), 339-353.

[22] Mačerinskienė, I. & Aleknavičiūtė, R. (2017). National intellectual capital influence on economic growth in the European Union countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *12*(4), 573-592.

[23] Mardani, A., Jusoh, A., Zavadskas, E.K., Kazemilari, M., Ungku, N.U.A. & Khalifah, Z. (2016). Application of Multiple Criteria Decision Making Techniques in Tourism and Hospitality Industry: a Systematic Review. *Transformations in Business & Economics*, 15(1)(37), 192-213.

[24] Meluzín, T., Balcerzak, A.P., Pietrzak, M. B., Zinecker, M., & Doubravský, K. (2018). The impact of rumours related to political and macroeconomic uncertainty on IPO success: evidence from a qualitative model. *Transformations in Business & Economics*, 17, 2(44), 148-169.

[25] Meluzín, T., Pietrzak, M. B., Balcerzak, A. P., Zinecker, M., Doubravský, K., & Dohnal, M. (2017). Rumours Related to Political Instability and their Impact on IPOs: The Use of Qualitative Modeling with Incomplete Knowledge. *Polish Journal of Management Studies*, 16(2), 171-187.

[26] Meluzín, T., Zinecker, M., Balcerzak, A.P., Doubravský, K., Pietrzak, M. B., & Dohnal, M. (2018), The timing of initial public offerings – non-numerical model based on qualitative trends. *Journal of Business Economics and Management*, 19(1), 63-79.

[27] Oprean C., Tănăsescu C. & Bucur A. (2017). A new proposal for efficiency quantification of capital markets in the context of complex non-linear dynamics and chaos. *Economic Research-Ekonomska Istraživanja*, 30(1), 1669-1692.

[28] Pietrzak. M. B. & Balcerzak, A. P. (2016). Socio-economic Sustainability in Poland. SEM Analysis at Regional Level. In T. Kliestik (Ed.). *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.).* Zilina: University of Zilina, 1704-1712.

[29] Pietrzak, M. B. & Balcerzak, A. P. (2016). A Spatial SAR Model in Evaluating Influence of Entrepreneurship and Investments on Unemployment in Poland. In M. Reiff & P. Gezik (Eds.). *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XVIII*. Vratna: Letra Interactive, 303-308.

[30] Pietrzak, M. B. & Balcerzak, A. P. (2017). A Regional Scale Analysis of Economic Convergence in Poland in the Years 2004-2012. In M. H. Bilgin and H. Danis, E. Demir, & U. Can (Eds.). *Regional Studies on Economic Growth, Financial Economics and Management. Proceedings of the 19th Eurasia Business and Economics Society*. Vol. 7. Springer International Publishing, 257-268.

[31] Pietrzak, M. B. & Balcerzak, A. B. (2017). Economic development of Polish voivodeships in the years 2010-2014. Application of taxonomic measure of development with entropy weights. In M. Papież and S. Śmiech (Eds.), *The 11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*. Cracow: Foundation of the Cracow University of Economics, 310-318.

[32] Pietrzak, M. B., Balcerzak, A. P., Gajdos, A., & Arendt, Ł (2017). Entrepreneurial environment at regional level: the case of Polish path towards sustainable socio-economic development. *Entrepreneurship and Sustainability Issues*, *5*(2). 190-203.

[33] Pilc, M. (2017). Cultural, political and economic roots of the labor market institutional framework in the OECD and post-socialist countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *12*(4), 713-731.

[34] Rogalska, E. (2017). Multivariate analysis of entrepreneurship in Poland at regional level. In T. Kliestik (Ed.). 17th 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, 4th – 5th October 2017. (Part V.). Zilina: University of Zilina, 2198-2202.

[35] Rogalska (2018). Entrepreneurship Conditions in Poland at NUTS 3 Level. Application of taxonomic measure of development based on median vector Weber. In M. Papież and S. Śmiech (Eds.), *The 12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings.* Cracow: Foundation of the Cracow University of Economics.

[36] Rollnik-Sadowska, E. & Dąbrowska, E. (2018). Cluster analysis of effectiveness of labour market policy in the European Union. *Oeconomia Copernicana*, *9*(1), 143-158.

[37] Šafránková, J. & Šikýř, M. (2017). Work expectations and potential employability of millennials and post-millennials on the Czech labor market. *Oeconomia Copernicana*, 8(4), 585-599.

[38] Shannon, C. E. & Weaver, W. (1947). *The mathematical theory of communication*. Urbana: The University of Illinois Press.

[39] Shemshadi, A., Shirazi, H., Toreihi, M. & Tarokh, M. J. (2011). A fuzzy VIKOR method for supplier selection based on entropy measure for objective weighting. *Expert Systems with Applications*, 38(10), 12160-12167.

[40] Simionescu, M., Lazányi, K., Sopková, G., Dobeš, K. & Balcerzak, A. P. (2017). Determinants of Economic Growth in V4 Countries and Romania. *Journal of Competitiveness*, 9(1), 103-113.

[41] Stankeviciene, J., Kraujaliene, L. & Vaiciukeviciute, A. (2017). Assessment of technology transfer office performance for value creation in higher education institutions. *Journal of Business Economics and Management*, *18*(6), 1063-1081.

[42] Tomovska Misoska, A., Dimitrova, M. & Mrsik, J. (2016). Drivers of entrepreneurial intentions among business students in Macedonia. *Economic Research-Ekonomska Istraživanja*, 29(1), 1062-1074.

[43] Wang, T. C. & Lee, H. D. (2009). Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. *Expert Systems with Applications*, 36(5), 8980-8985.

[44] Vavrek R., Adamisin P. & Kotulic R. (2017). Multi-criteria evaluation of municipalities in Slovakia – case study in selected districts. *Polish Journal of Management Studies*, 16(2), 290-301.

[45] Zygmunt, J. (2018). Entrepreneurial activity drivers in the transition economies. Evidence from the Visegrad countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *13*(1), 89-103.

Technological Progress at the Sectoral Level: the Sato Production Function Approach

Lenka Roubalová[1], David Hampel [2], Lenka Viskotová[3]

Abstract. The main aim of this paper is to reveal the relationship between the economic output, labour and capital, and to determine technological progress type in the four sectors of the manufacturing industry in selected Central European countries during the period 1995–2015. The methodology is based on the Sato production function which is a not frequently used case of the non-CES production function. For this purpose, the original Sato specification is modified by adding the time variable that allows us to observe how the relationship between the production and the input factors changes over time. The Nonlinear Least Squares method with an appropriate initial values setting is used to estimate its parameters. Based on these parameters the isoquant maps are depicted and the borders of the economically efficient region are determined. Within that area, the isoquant slope development described by the values of the marginal rate of technical substitution is used to identify the type of the technological progress in the investigated sectors for each country. In spite of some individual characteristics the results also prove the common features of the development.

Keywords: MRTS, Nonlinear Least Squares, Sato production function, technological progress

JEL classification: C51, D24, O47
AMS classification: 62J02, 91B38

1 Introduction

The history of production function (PF) is associated with the classical economics. Since then, various specifications and ways how to use it have developed, see [4]. Among other things, the production functions are used as a tool of technological progress (TP) measuring. This approach is associated with Keynesians or rather Neokeynesians. Famous names like Solow or Hicks should be mentioned here [3], [9]. Another author [5] focused on PF and Data Envelopment Analysis to evaluate the technological progress of the smart grid or [19] investigated the TP via Cobb-Douglas PF and Constant Elasticity of Substitution (CES) PF.

In the long run, the concept of technological progress assumes the production function is not fixed, but it reflects the changes in a production process. In particular, technological progress decreases the quantity of inputs needed to achieve given level of output. Economic theory classifies technological progress into three types based on Hicks [9]. For the neutral technological progress the marginal rate of technical substitution (MRTS) is kept unchanged along the ray from the origin. It means that the isoquants corresponding to the given level of output shift inward but the isoquant slope keeps unchanged. If MRTS of capital for labour is increasing, the isoquants in the LK-plane become steeper, because the absolute value of the isoquants slope is increasing. Marginal product of labour (MPL) increases at the faster pace than marginal product of capital (MPK), that is typical for capital-saving TP. Likewise, if the time series of the MRTS is decreasing, the isoquants become flatter, because the absolute value of the isoquants slope is decreasing. It indicates the MPL increases at the slower pace than the MPK. Such kind of development characterizes labour-saving TP [3].

The original version of the Sato function is based on the assumption that its two parameters a, b, are constant. This function exhibits constant returns to scale and it is linearly homogeneous. Initially, it was proposed by Sato [16], [15] who based his work on the modification of the CES function, that can be extended to the n-input case and to the variable elasticity substitution case, if the elasticity of substitution depends on the output level. For the Sato function, the elasticity of substitution is constant along an isoquant, but it varies across the isoquants as the output varies. It is possible when an isoquant in LK-plane is of Cobb-Douglas shape or CES shape, but the another one can be non-homotethic CES. Therefore, the Sato function contains rather a kind of non-CES production functions.

[1]Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, xkravack@node.mendelu.cz
[2]Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, david.hampel.uso@mendelu.cz
[3]Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, lenka.viskotova@mendelu.cz

The main objective of this research is to reveal the relationship between the economic output, and multiple inputs, labour and capital, via the time augmented Sato PF proposed here. It is the way how to determine the technological progress type in the four sectors of the manufacturing industry in the Czech Republic (CZ), Slovakia (SK), Austria (AT) and Germany (D) during the period 1995–2015.

2 Material and Methods

The Sato production function is given by the equation

$$Y = L^2K^2/(aL^3 + bK^3), \tag{1}$$

where Y represents production, K, L are input factors, capital and labour, and a, b are its parameters, where $a > 0$, $b > 0$. To obtain an adequate tool to capture the development over time, we propose using of the augmented Sato function. Parameters a, b, are supposed to be a function of time $a_t = a_0 + a_1t$, $b_t = b_0 + b_1t$, where t means time (usually $t = 1, ..., T$). Then, both parameters can vary over time which implies the production function and the marginal product of the individual inputs vary as well. Such time augmented function is given by the equation

$$Y = L^2K^2/(a_0L^3 + a_1tL^3 + b_0K^3 + b_1tK^3), \tag{2}$$

where a_0, a_1, b_0, b_1 are its parameters (while a and b are decomposed into a_0, a_1, b_0, b_1). This modification makes the analysis dynamic and it allows us to observe the development of the relationship between production and the input parameters continuously, year by year over the whole period.

The NLS technique is used to estimate the parametres for all four sectors of all four countries. As the function is non-linear in parameters, this method appears to be the most appropriate. Another way how to estimate the parameters of the function non-linear in parameters is to linearize it and then use linear techniques, e.g. OLS metod. However, obtained results can differ from the parameters estimated by NLS, because using of the linearization and OLS estimation can lead to the model with the residual variance that is not minimized in fact. As criteria to select the best estimated model we use R^2 coefficient of determination and Theil's U [17] to compare models from the point of view of the accuracy of the forecast. Theil's U has a minimum of 0, the more accurate the forecast is, the lower is the Theil's U value. The nonlinear model estimation is computed in iterations using numerical methods, see [10] or [12]. To run this procedure correctly it is required to set suitable initial values of all parameters. Therefore, the total count of the iterations leading to the final model estimation is also the criterion we evaluate. Although the end of the estimating procedure should not be based on the number of iterations, it is assumed that the less iteration is needed, the more stable the model is. Models with statistically significant parameters ($\alpha = 0.05$) are preferred, but slightly insignificant parameter should not be excluded as the main aim is to find the best estimation of all Sato production function parameters and keep it complete.

Based on the estimated parameters we depict isoquants in LK-plane to define the economically efficient region of production. In that area the isoquants are downward sloping and the production costs leading to the given production of output are minimized. The suitable part of each isoquant curve is located between the point where the slope of the tangent line is either zero or equal to ∞. These points satisfy the equations $MPL = 0$, $MPK = 0$, where MPL and MPK are marginal product of labour and marginal product of capital which are obtained as the partial derivatives of the production function.

Generally, the curves representing borders of the economically efficient region can be derived, see [3]. In the case of the Sato function the borders are straight lines going through the origin of the LK-plane, given by the equations

$$K = \sqrt[3]{\frac{2a}{b}}L, \qquad K = \sqrt[3]{\frac{a}{2b}}L. \tag{3}$$

Then, the MRTS of capital for labour can be calculated to identify the technological progress type in each sector. The MRTS is given by the ratio of MPL and MPK. We use an average production of each sector of each country and the estimated parameters of each individual sector to obtain the isoquant of the average production. Then, a straight line from the origin representing a ray identifying the tangent point of the isoquant and its tangent can be constructed and the MRTS (the negative of the isoquant slope) can be calculated for the selected isoquant of each year.

The dataset consists of the annual time series for the period from 1995–2015 ($T = 21$) sourced from the EU KLEMS database (EU KLEMS, 2018) including data of four sectors of the manufacturing industry: processing of food, beverages and tobacco (FOOD); wood and paper products, printing and reproduction of recorded media (WOOD); basic metals and fabricated metal products, except machinery and equipment (METAL); machinery and equipment (MACHINERY). As variables we use gross output at current basic prices for Y, the nominal gross fixed capital formation for K, both in millions of the national currencies and the total sum of hours worked for L.

3 Results

The results for individual countries are summarized below in Tab.1–Tab 4. All the estimates meet the condition of $a > 0, b > 0$. For each sector of each country the best estimated (from the view-point of R^2 coefficient and Theil's U value) model is presented here. Based on the estimated parameters the isoquant curves can be depicted in the LK-plane for each year. In all cases we revealed the isoquants shift inward and the isoquant slope changes. It means that for the same production less inputs are needed, so the marginal productivity of inputs increases. As an example of the development, see results for the Czech food processing sector presented in Fig. 1.

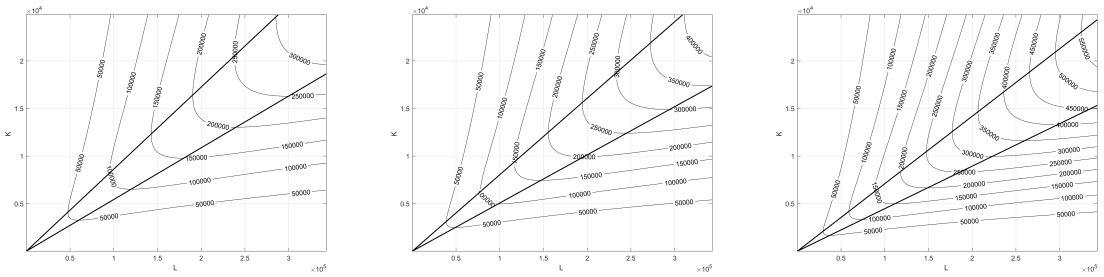


Figure 1 LK-plane for the Czech food sector in years 1995, 2005 and 2015. The straight lines define the economically efficient region.

Looking at the results for SK in Tab. 2, we can see the highest Theil's U for all sectors in comparison with the other countries. Further analysis shows that these models are not able to describe the relationship between production and inputs properly as its development is not consistent during the whole period. By computing the borders of the economically efficient region and investigating the intersection of the isoquant and the economically efficient region for particular years we reveal that there is an empty intersection in some cases, especially for SK. The production is not efficient in these years so this periods are not taken into account for further analysis based on MRTS. Especially, this is the issue of the food processing sector (years 2009–2015), the wood processing sector (years 2014, 2015) and the metal processing sector (years 1995, 1996) in Slovakia and wood processing sector (years 2013, 2014, 2015) in the Czech Republic.

	FOOD	**WOOD**	**METAL**	**MACHINERY**
$\widehat{a_0}$	$2.248 \cdot 10^{-3}$	$4.672 \cdot 10^{-3}$	$4.106 \cdot 10^{-3}$	$5.456 \cdot 10^{-3}$
$\widehat{a_1}$	$-6.849 \cdot 10^{-5}$	$-1.663 \cdot 10^{-4}$	$-1.402 \cdot 10^{-4}$	$-1.975 \cdot 10^{-4}$
$\widehat{b_0}$	7.447	8.784	4.501	10.116
$\widehat{b_1}$	$-9.120 \cdot 10^{-2}$	$3.252 \cdot 10^{-2}$	$1.296 \cdot 10^{-1}$	$-2.392 \cdot 10^{-1}$
R^2	45.407 %	93.520 %	80.087 %	95.458 %
U	2.107	1.542	1.268	0.722

Table 1 Estimated parameters of the Sato function for the Czech Republic, coefficient of determination, Theil's U.

	FOOD	**WOOD**	**METAL**	**MACHINERY**
$\widehat{a_0}$	$1.108 \cdot 10^{-4}$	$1.998 \cdot 10^{-4}$	$5.172 \cdot 10^{-5}$	$1.104 \cdot 10^{-4}$
$\widehat{a_1}$	$-5.772 \cdot 10^{-6}$	$-6.875 \cdot 10^{-6}$	$1.223 \cdot 10^{-6}$	$3.700 \cdot 10^{-6}$
$\widehat{b_0}$	6964.970	3481.240	4542.11	8775.25
$\widehat{b_1}$	-83.750	-57.860	-143.514	-355.239
R^2	77.096 %	36.826 %	69.223 %	91.883 %
U	2.410	3.473	2.394	1.235

Table 2 Estimated parameters of the Sato function for Slovakia, coefficient of determination, Theil's U.

	FOOD	**WOOD**	**METAL**	**MACHINERY**
$\widehat{a_0}$	$7.909 \cdot 10^{-5}$	$1.654 \cdot 10^{-4}$	$1.651 \cdot 10^{-4}$	$2.454 \cdot 10^{-4}$
$\widehat{a_1}$	$-6.206 \cdot 10^{-7}$	$-2.925 \cdot 10^{-6}$	$-3.444 \cdot 10^{-6}$	$-2.444 \cdot 10^{-6}$
$\widehat{b_0}$	1332.100	1150.030	946.860	640.858
$\widehat{b_1}$	-28.097	-25.477	-15.193	-11.046
R^2	95.627 %	50.355 %	81.249 %	92.540 %
U	0.768	1.068	0.955	0.771

Table 3 Estimated parameters of the Sato function for Germany, coefficient of determination, Theil's U.

	FOOD	**WOOD**	**METAL**	**MACHINERY**
$\widehat{a_0}$	$1.237 \cdot 10^{-4}$	$2.586 \cdot 10^{-4}$	$4.181 \cdot 10^{-4}$	$3.031 \cdot 10^{-4}$
$\widehat{a_1}$	$-2.322 \cdot 10^{-6}$	$-7.773 \cdot 10^{-6}$	$-1.453 \cdot 10^{-5}$	$-1.621 \cdot 10^{-6}$
$\widehat{b_0}$	1516.4	786.037	545.004	670.735
$\widehat{b_1}$	-47.465	-14.917	-6.676	-16.493
R^2	97.372 %	91.912 %	94.918 %	93.992 %
U	0.731	0.752	0.691	0.729

Table 4 Estimated parameters of the Sato function for Austria, coefficient of determination, Theil's U.

Further, the MRTS trend is calculated in all years with respect to the isoquant of the same production, see Fig. 2 and Fig 3. Looking at the issue from the point of view of countries, we can summarize the results as follows. For CZ in all sectors capital-saving TP is identified. The MRTS trend is increasing, the MPL increases at the faster pace than the MPK so the labour force is more productive and the technological change goes through the labour force, e.g. increasing skill level and qualification.

In the case of SK and D, mainly labour-saving TP is identified. The two sectors for which capital-saving TP is detected are described in Discussion. These results suggest the technological change is driven by the improvement of the capital equipment, because the MPK increases at the faster pace than the MPL. Meanwhile in the case of Germany the MRTS development dynamics is very similar for all sectors, for Slovakia a sharp rise of MRTS for the food processing sector and sharp decline for the metal processing sector can be observed.

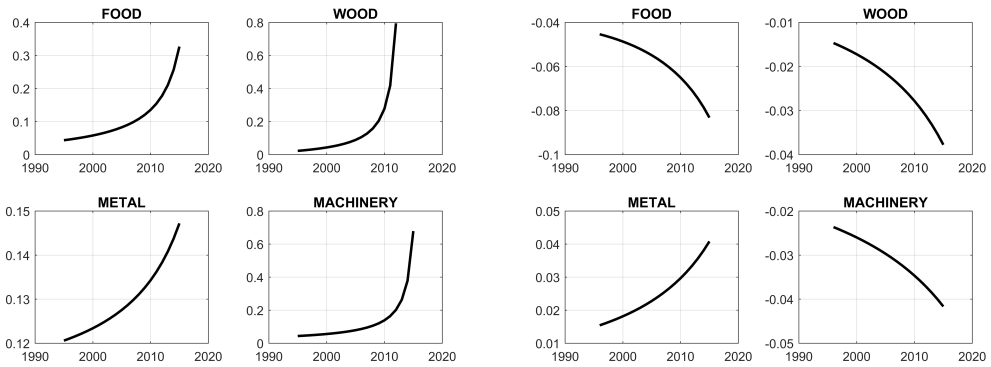


Figure 2 The graphs describing the MRTS trend during the observed period in each sector (MRTS on the vertical axis, time *t* on the horizontal axis). The Czech Republic on the left-hand side, Germany on the right-hand side.

For AT the results are diverse. On the first hand, the development of food processing and machinery industry is very similar indicating that technological progress was driven by the continuously increasing marginal productivity of capital. On the other hand, increasing MRTS can be observed for the wood and the metal processing sector, but the dynamics of these cases differs. Firstly, the MRTS growth is very slow then, after 2010 it increases rapidly. In other words, the MPL growth was much more faster than the MPK growth in these two sectors.

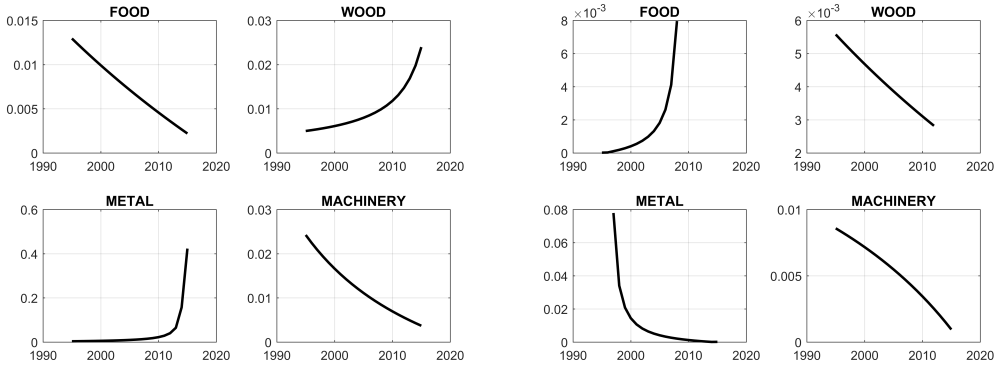


Figure 3 The graphs describing the MRTS trend during the observed period in each sector (MRTS on the vertical axis, time *t* on the horizontal axis). Austria on the left-hand side, Slovakia on the right-hand side.

4 Discussion

For CZ, the results show that technological progress in all selected sectors of manufacturing industry is driven by the labour force performance. There is an evidence of upgrading knowledge in manufacturing industry for the period 2007–2012, see [8]. These results are partly supported by the paper [18] focused on innovations in manufacturing industry in CZ between 1999–2008. It was proved that multinational companies lead the innovation and the most hampering factors of the innovations are the innovation costs and lack of information and lack of skilled workers. As we proved in [14], also labour costs are one of the aspects significantly affecting labour productivity in the European countries.

Based on our results we can also conclude, that this kind of progress is characteristic for metal processing industry in general. For D, this is the only sector where capital-saving progess is detected. For SK and D labour-saving technological progress was found. These findings are in accordance with [6] arguing that machinery industry is the largest and the most innovative industry in Germany and it is considered to be a leader of the technological development. Authors state that mainly automotive manufacturers have been substantially investing in modernization, capacity expansion and new technologies, mainly since 2010. Moreover, publication [6] explains the demand for robotics is also being felt for example in food sector.

Based on the growth accounting approach, in Baran [2] states that capital was the most significant growth input during 1995–2010 in Central Europe. Moreover, her results show, that at the country level, capital contribution to the growth was the most distinct in Slovakia. Similarly, our results revealed that the marginal productivity of capital grows at the faster pace than the marginal productivity of labour for Slovakia and Germany. The food processing industry is an only exception in SK for which capital-saving TP is identified but only for the period 1995–2008. Further analysis proved that there are no economically effective combinations of inputs corresponding to the isoquants based on the estimated parameters for period 2009–2015. The reason may be found in the continuous decrease of capital for the food processing industry after 2007 followed by the most sinificant sharp decline in 2012. Such completely different developement probably causes that the model does not fit for all this period. In [13] we can find the statement that both years, 2009 and 2012, correspond with the trough phase of the crisis.

Our results proved the time augmented Sato PF is suitable for analysis at the sectoral level. With respect to the estimated model we confirmed the time can be considerated to be an important input factor and this augmented version is an adequate tool of measuring the development of the output and input relationship. Despite the fact the Sato PF is not frequently used tool in economics, the research [11] can be given as an example of using this function. Their results confirm that the Sato production function is suitable for analysis either at the aggregate level or in the microeconomic field.

5 Conclusion

We proved it is possible to capture the marginal productivity development (with use of MRTS) via the augmented Sato function which is suitable for the description of the reationship of Y, K, L. The results also show the time variable is an important input factor. Moreover, we can say that in the case of some SK and CZ sectors, there are periods for which the production is not efficient. In general, it can be concluded that technological progress is driven by labour force quality in the Czech manufacturing industry and by the technical advance in capital equipment in Slovakia and Germany. For Austria the results are heterogeneous. However, this should not be

generalized absolutelly. It is probable that at the aggregate level the development can be different due to the other sectors not included in this analysis. Further research may be focused on country level to reveal common features of labour, capital and output relationship development. It might be compared with the results obtained via analysis focused on sectoral level. Another way could be the comparison with the results obtained via Cobb-Douglas production function estimates.

Acknowledgements

This research is supported by the doctoral grant No. $PEF-DP-2018029$ of the Grant Agency IGA PEF MENDELU.

References

[1] Balcar, J., Jankov, L., Filipov L. (2014). What General Competences are Required frm the Czech Labour Force? *Prague Economic Papers*, 2, 250–265.

[2] Baran, K., A. (2013). The Determinants of Economic Growth in Hungary, Poland, Slovakia and the Czech Republic During the Years 1995-2010. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 8, 7–26.

[3] Besanko, D., Braeutigam, R. (2007). *Microeconomics.* New York: John Wiley & Sons Inc.

[4] Blaug, M. (1985). *Economic Theory in Retrospect*. Cambridge: Cambridge University Press.

[5] Dong, H., Zheng, Y. (2014). Measuring Technological Progress of Smart Grid Based on Production Function Approach. *Mathematical Problems in Engineering*, 2014, 1–9.

[6] Germany Trade and Invest. (2017). The Machinery and Equpiment Industry in Germany. *Industry overview*. [Online]. `https://www.gtai.de/GTAI/Content/EN/Invest/_SharedDocs/Downloads/GTAI/Industry-overviews/industry-overview-machinery-equipment-en.pdf?v=12`. [Accessed: 18 April 2018].

[7] EU KLEMS. (2017). *EU KLEMS Growth and Productivity Accounts: Statistical Module, ESA 2010 and ISIC Rev. 4 industry classification*. [Online]. `http://www.euklems.net/`. [Accessed: 18 April 2018].

[8] European Commision. (2014). *Research an Innovation Performance in Czech Republic Country profile*. [Online]. Available at: `https://ec.europa.eu/research/innovation-union/pdf/state-of-the-union/2014/countries/czech_republic.pdf`. [Accessed: 18 April 2018].

[9] Hicks, J., R. (1932). *The Theory of Wages*. London: Macmillan.

[10] Levenberg, K. (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2, 164-168.

[11] Makin, J. A., Strong, S. (2013). New Measures of Factor Productivity in Australia: a Sato approach. *Applied Economics Australia*, 45, 2413–2422.

[12] Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameter. *Journal of the Society for Industrial and Applied Mathematics*, 11, 431–441.

[13] Morvay, K. et al. (2013). *Hospodársky vývoj Slovenska v roku 2012 a výhlad do roku 2014*. Bratislava. Ekonomický ústav Slovenskej akadémie vied.

[14] Roubalová, L., Vaněk, T., Hampel, D. (2017). On the Relation of Labour Productivity, Costs and Unemployment in the EU Countries. *Proceedings of the 35th International Conference Mathematical Methods in Economics* (pp. 632–637)

[15] Sato, R. (1964). Diminishing Returns and Linear Homogeneity: comment. *The American Economic Review*, 54, 744–745.

[16] Sato, R. (1975). The Most General Class of CES Functions. *Econometrica*, 43: 999-1003.

[17] Theil, H. (1966). *Applied Economic Forecasting*, Amsterdam: North-Holland.

[18] Vokoun, M. (2016). Innovation Behaviour of Firms in a Small Open Economy: the Case of the Czech Manufacturing Industry. Austrian Institute for Economic Research, 43, 111–139.

[19] Voosholz, F. (2014). The Influence of Different Production Functions on Modeling Resource Extraction and Economic Growth. *CAWM Discussion Paper*, 72, 1–16.

Optimal Loan Performance Management via Stochastic Programming

Tomáš Rusý[1]

Abstract. Companies, whose business consists of providing loans to clients, must make numerous decisions during the life of their loans. They need to make an offer to the client and decide on how to finance such a loan. During these decisions, the profitability as well as all the credit and interest rate risks must be taken into account. We develop a non-linear, decision dependent randomness stochastic program which allows us to find the optimal sequence of decisions. We also discuss specific peculiarities of this loan type such as loan prepayments and client defaults.

Keywords: stochastic programming, decision dependent randomness, consumer loan.

JEL classification: C610
AMS classification: 90-04, 90C15, 90B50

1 Introduction

Stochastic programming is a quickly developing area of mathematical optimization, which finds its application in many different fields. Especially the area of financial planning and control has seen plenty of implementations of stochastic programming models. One such an application arose in 1986, when Kusy and Ziemba published their famous paper dealing with a bank asset-liability model [9]. Two years later, Dempster and Ireland [2] have introduced a different model where they focused on immunisation of liabilities and hence they for the first time took into account risks, which are associated with financial problems. Between other contributions, Carino et al. [1] presented a very first asset-liability model for an insurance company, which was lately followed by many others. We should mention at least some focused on insurance companies [7, 10], and pension funds [3, 6], as these two were the most popular types of financial companies where stochastic programming was used.

These works however did not exhausted the potential of applications of stochastic programming in finance. Dupačová and Kopa [4, 5] have proposed to investigate the robustness of program's solution by scenario contamination. Other possibility was to analyse the effect of new, modern risk constraints on the optimal solution; for example Kopa et al. [8] have applied a second-order stochastic dominance constraint. An important aspect of the modelling exercise is to focus on the accuracy of the model formulation. Vitali et al. [11] have paid a lot of attention to this part, as they aimed to cover all possibilities, how an insurance company could invest. They also focused on the credibility of cash-flow dynamic in the company and to the scenario generation procedure.

In our work, we focus on a stochastic programming formulation of a loan management problem, where we aim to maximise the value of a loan for the company. The program describes a whole life-cycle of a loan, which begins once client expresses his desire to borrow money and ends at the agreed time of maturity. We implement the impact of a client default and a loan repayment onto the value of a loan. In this short paper, in Section 2, we describe formulation of the model and our approach of generating scenarios. Thereafter, in Section 3, we show results of this program for one parameters' setting. There, we also talk about possible extensions of the program.

2 Model Formulation

We assume that a client comes to the company and wants to borrow N_0 amount of money for the period of T months. Such a consumer loan is repaid regularly each month $t, t = 1, \ldots, T$. In our program, we model payments to be exchanged at each such a month t. Our multi-stage stochastic program will consists of $K+1 \leq T+1$ stages, $0 = t_0 < \ldots < t_K = T$. In these times, scenarios of random quantities will be observed and the company will be allowed to make decisions.

2.1 Interest Rate Decision, Instalment and Principal

First, we derive the value of the instalment, which is constant over the whole period. The (annual, monthly compounded) interest rate, which is offered by the company to the client is denoted r and it is a decision variable. From there, we can derive the monthly instalment π, which client pays every month, as in (1). Next, we denote $N_t, t = 1, \ldots, T$, the principal which stays on the account after the instalment at time t. We can calculate the

[1]Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75, Praha 8, rusy@karlin.mff.cuni.cz

interest credited as from month $t-1$ to t as $N_{t-1} \cdot (r/12)$, while the amortization equals $\pi - N_{t-1} \cdot (r/12)$. Moreover, we can easily figure out value of N_1 and by induction also other principal amounts $N_t, t = 1, \ldots, T$.

$$N_0 = \sum_{t=1}^{T} \frac{\pi}{(1+r/12)^t} \quad \Rightarrow \quad \pi = N_0(r/12)\Big(1-(1+r/12)^{-T}\Big)^{-1} \tag{1}$$

$$N_t = N_0(1+r/12)^t \frac{(1+r/12)^{T-t}-1}{(1+r/12)^T-1}, \qquad t = 1, \ldots, T. \tag{2}$$

Obviously, the principal behaves in the way we calculated only in the case when the loan is repaid as agreed at the beginning of the contract. In case of a default of the client or a full/partial repayment of the loan, the evolution would be different. This will be discussed later at the end of Section 2.2.

2.2 Scenarios of Random Elements

This section is devoted to the description of all random elements which affect the life of a loan. We will talk about their implementation into the model and shortly also about the scenario generation. Randomness influencing our consumer loan can be divided into four parts. We will discuss all of them.

First, once the company offers client the loan with certain interest rate, he has the opportunity to either accept or reject the loan. This random effect gets more complicated by the fact that client's decision depends on the value of interest rate he is offered. The probability that the client closes the contract will be denoted by $p(r)$. This function could be client specific and his other characteristics can affect it too. We think of such a function to be estimated by logistic regression, where one of the regressors would be the offered rate to the client. Hence in our program, we employed a function

$$p(r) = \frac{\exp\{b_0 + b_1 r\}}{1+\exp\{b_0 + b_1 r\}} = \frac{1}{1+\exp\{-(b_0 + b_1 r)\}}, \tag{3}$$

where b_0 and b_1 are some (client-dependent) parameters, with $b_1 \leq 0$, as increase in offered interest rate should decrease client's probability of accepting the loan. As this decision comes before issuing the loan, we can express the objective function (which will be discussed later) as a sum of two parts — the value of the program when the loan is and is not closed. These would be then weighted by the corresponding probabilities $p(r)$ and $1-p(r)$.

Next, we will talk about interest rates which express the cost of money at the market. We denote y_t^τ the annualized, risk free interest rate with time-to-maturity τ at time t. For $t > 0$, this quantity is random. We also denote m_t^τ the rates, for which the company can borrow from a bank. We define: $m_t^\tau = y_t^\tau + m(\tau)$.

Quantity $m(\tau)$ represents the differences between the risk free rates and the rates a bank charges the company. Hence it can be interpreted as the mark-up of the bank. In the model, we generate scenarios of interest rate from the Hull - White model, which is calibrated to the observed market data. We generate regular scenario tree often seen in multi-stage stochastic programming — one where in each stage every node has the same number of successors. Such a tree has all scenarios equiprobable. We will denote S_i the set of nodes of the interest-rate tree in a decision stage t_i and $a_i(s_j)$ the time t_i ancestor of a node $s_j \in S_j, i < j$. We will denote $y_{t_i}^\tau(s_i)$ the risk free interest rate and $m_{t_i}^\tau(s_i) = y_{t_i}^\tau(s_i) + m(\tau)$ the rates for the company at the scenario node $s_i \in S_i$.

Let us now move to the final two types of randomness, which enter the program. These are loan prepayments and client defaults. In general, such events are something the company cannot really affect so we take them as exogenous. First, let us describe what we want to take into account by including these effects into the model.

Client defaults are generally considered as the biggest risk factor affecting the profitability of a loan. That is simply because it can happen that the client becomes unable to fulfill his commitments. Under such an event, the company loses not only the interest rate charged to the client, but also a part of the principal. Such a proportion is called *loss given default*, and we include it into the model as a fixed parameter, denoted as lgd. If client defaults at time t_k, the company is modelled to receive $(1-\mathrm{lgd})N_{t_{k-1}}$ as the recovered part of the loan. Scenarios of client defaults will be added to the model. Thereafter, we assign them probabilities so they match our initial assumption about the hazard rate $h(t_k)$ — the probability of default at time t_k given that the loan survived up to time t_{k-1}.

Prepayments can be considered as the complete opposite of client defaults, as it means client repays more money than as agreeed in the contract. Usually, such a prepayment comes from one of the two following reasons:

- Client has spare money which he can afford to use for loan repayment.
- Client finds an opportunity how to borrow money cheaper and he refinances the loan.

The first reason is kind of completely random to us and it is not really connected to other decision variables or random quantities in the problem. On the other hand, the second reason is closely related to interest rate, as when it

decreases, companies offer lower rates which makes the probability of refinancing the loan higher. Similarly with client defaults, we add scenarios of loan prepayments into the problem and assign them some probabilities in a way we match our assumptions on hazard rate $g(t_k, s_k)$ at time t_k of loan prepayment at the node $s_k \in S_k$. Such a hazard rate is node specific because of its connection to interest rate. We should clarify that by loan (p)repayment, we mean only full repayment of the loan.

Next, let us discuss how we implement the scenarios of loan prepayments and client defaults into the program. First, we should realise that we have $|S_K|$ interest rate scenarios from the initial decision period to the maturity of the loan. Take one scenario as fixed, then at each node/stage, loan prepayment or client default can occur. We should also note that during the life of the loan, only one client default or only one loan repayment can take place. Hence each program scenario can be defined as a pair of interest rate scenario and the event specification. Events in this case are loan prepayments/client defaults at any stage $t_k \neq t_0$. We should also mention that loan prepayment at time $t_K = T$ corresponds to the loan being repaid as initially agreed.

Mathematically speaking, we define set E as a set of all possible events. For event $e \in E$, we define $t(e)$ to be the time when the event occurs, $d(e)$ an indicator function taking 1 when the event is client default and 0 when the event is full repayment. So for example for event of client default at the final stage T, we have $t(e) = T, d(e) = 1$. Scenario of a program is then uniquely defined by a pair $(s, e), s \in S_K, e \in E$.

What is important to realise under this parametrisation is that at time t_k, we cannot distinguish between two scenarios $(s, e_1), (s, e_2), s \in S_K, e_1, e_2 \in E$ such that $t_k < \min\{t(e_1), t(e_2)\}$. That is simply from the reason that by the time t_k, we do not observe which event from E has happened. Such a property will lead us to the inclusion of non-anticipativity constraints into the program.

In the final part of this section, we will comment on how we calculate probabilities of scenarios $p(s, e), s \in S_K, e \in E$. These are obtained iteratively by multiplication of marginal (transition) probabilities in each node. Before the time when event $e \in E$ takes place we make use of hazard rates$f(t_k)$ and $g(t_k, s_k)$, otherwise, probabilities are distributed equally depending on branching of the interest rate tree. The formula looks as follows:

$$p(s,e) = \prod_{k=1}^{K} \Big(\mathbb{I}_{[t(e)=t_k]} d(e) h(t_k) + \mathbb{I}_{[t(e)=t_k]} \big(1 - d(e)\big) g\big(t_k, a_k(s)\big) + \mathbb{I}_{[t(e)<t_k]} \big(1 - h(t_k) - g(t_k, a_k(s))\big) + \mathbb{I}_{[t(e)>t_k]} \Big) \big(|S_{k-1}|/|S_k|\big), \quad s \in S_K, e \in E.$$

We require that $\forall s \in S_K$ and $\forall k \in 1, \ldots, K$ it holds $h(t_k) \geq 0, g(t_k, a_k(s)) \geq 0, \big(1 - h(t_k) - g(t_k, a_k(s))\big) \geq 0$. Moreover, in the last stage, we must have $g(t_K, s) = 1 - h(t_K), \forall s \in S_K$. Under such conditions, scenarios' probabilities are non-negative and sum up to one.

2.3 Cost of Financing the Loan

Another aspect, which needs to be considered is the cost of financing such a loan. The company can decide how to form the liability side to obtain funds to provide the loan. We will give the company a whole panoply of options. Consider now two time instances t_i and $t_j, t_i < t_j$ of the program, where possibly $t_j > T_K$. In each node of every scenario $(s_i, e), s_i \in S_i, e \in E, i \in 1, \ldots, K$, the company will have two possibilities how to borrow money from a bank. It could choose either to borrow from t_i to t_j such that the loan is repaid continuously at each month $t, t_i < t \leq t_j$ or such that all the money is repaid at once at time t_j. We will denote such amounts $u_{t_i,t_j}(s_i, e)$ and $v_{t_i,t_j}(s_i, e)$ respectively. We can calculate the amount $u_{t_i,t_j}(t; s_i, e)$ repaid at time $t, t_i < t \leq t_j$ from a loan $u_{t_i,t_j}(s_i, e)$ as

$$u_{t_i,t_j}(t; s_i, e) = \frac{u_{t_i,t_j}(s_i, e)}{\sum_{\tau=1}^{t_j - t_i} \big(1 + m_{t_i}^{\tau}(s_i)/12\big)^{-\tau}}, \qquad t = t_i + 1, \ldots, t_j. \tag{4}$$

On the contrary, at time t_j, the company pays back $v_{t_i,t_j}(t_j; s_i, e)$ such that

$$v_{t_i,t_j}(t_j; s_i, e) = v_{t_i,t_j}(s_i, e) \Big(1 + m_{t_i}^{t_j - t_i}(s_i)/12\Big)^{t_j - t_i}. \tag{5}$$

We also make it possible to the company to invest spare money and gain some interest. We denote $w_{t_i,t_j}(s_i, e)$ the amount of money lend to others for the market risk free yield $y_{t_i}^{t_j - t_i}(s_i)$. This money will be repaid at time t_j, as the company would receive amount $w_{t_i,t_j}(t_j; s_i, e)$:

$$w_{t_i,t_j}(t_j; s_i, e) = w_{t_i,t_j} \Big(1 + y_{t_i}^{t_j - t_i}(s_i)/12\Big)^{t_j - t_i}. \tag{6}$$

Summing all the equations up, we can express both the amount of the company's income (I) and expenditures (J) at time instance $t_k, t_k = 1, \dots, T$ from third party providers (excluding the client). These have the following form:

$$I_{t_k}(s_k, e) = \sum_{t_i : t_i < t_k} w_{t_i,t_k}(t_k; a_i(s_k), e) + \sum_{t_j : t_k < t_j} u_{t_k,t_j}(s_k, e) + \sum_{t_j : t_k < t_j} v_{t_k,t_j}(s_k, e), \tag{7}$$

$$J_{t_k}(s_k, e) = \sum_{t_j : t_k < t_j} w_{t_k,t_j}(s_k, e) + \sum_{\substack{t, t_i, t_j : \\ t_i \le t_{k-1} < t \le t_k \le t_j}} u_{t_i,t_j}(t; a_i(s_k), e) + \sum_{t_i : t_i < t_k} v_{t_i,t_k}(t_k; a_i(s_k), e). \tag{8}$$

2.4 Constraints and Objective Function

In the following section, we will go through all the constraints implemented in the model. Let us denote $B_{t_k}(s_k, e)$ the amount of money the company has on its account immediately after time t_k in scenario (s_k, e) and C_{t_k} the company's operating costs of such a loan from time t_k to t_{k+1}. In the formulation of the equation for the company's cash account, we need to be aware of the time and type of the event $e \in E$. That information defines the cash-flow the company receives from the client.

$$\begin{aligned} B_{t_k}(s_k, e) = {} & B_{t_{k-1}}(a_{k-1}(s_k), e) - C_{t_k} + I_{t_k}(s_k, e) - J_{t_k}(s_k, e) - \mathbb{I}_{[k=0]} N_0 \\ & + \mathbb{I}_{[t_k < t(e)]}(t_k - t_{k-1})\pi + \mathbb{I}_{[t_k = t(e) \& [d(e)=1]]} \mathrm{lgd} \cdot N_{t_{k-1}} + \mathbb{I}_{[t_k = t(e) \& [d(e)=0]]} \left((t_k - t_{k-1})\pi + N_{t_k}\right), \end{aligned} \tag{9}$$

for $k = 0, \dots, K$, where $B_{-1} = 0$, and also $C_{t_K} = 0$. The relationship on the first line of (9) expresses the initial exchange of notional and the cash-flows between the bank and the company. On the second line, the amount of funds the company receives from the client in different stages under the scenario e are described.

The definition of company's cash account brings us to a very natural survival condition such that the company cash account must not be lower than 0. We require it only in stages from $0, \dots, K-1$ as in the last stage, the loan is concluded and we look at the final balance, its performance thorough its life and asses its profitability. We require

$$B_{t_k}(s_k, e) \ge 0, \qquad e \in E, s_k \in S_k, k = 0, \dots, K-1.$$

Also at time T, the loan is matured, and the company is not allowed to lend or borrow money. We impose:

$$u_{t_K,t_j}(s_K, e) = 0, \qquad v_{t_K,t_j}(s_K, e) = 0, \qquad w_{t_K,t_j}(s_K, e) = 0, \qquad \forall t_j \ge t_K, s_K \in S_K, e \in E.$$

Next, we move to the cash-flows which take place between decision stages. The company has to make sure it has enough money to cover its expenditures up to the next decision stage. Such a constraint can be implemented only by checking whether the company has enough funds in the month before the next decision stage as all the cash-flows in between the decision stages are the same every month. The constraint looks as follows:

$$0 \le B_{t_k}(s_k, e) + \mathbb{I}_{[t_k < t(e)]}(t_{k+1} - t_k - 1)\pi - \sum_{\substack{t, t_i, t_j : \\ t_i \le t_k < t < t_{k+1} \le t_j}} u_{t_i,t_j}(t; a_i(s_k), e), k = 0, \dots, K-1. \tag{10}$$

The next step is to express the value of a loan in each scenario. Such a value is calculated as the discounted cash-flows from loans running at the given scenario. Let us denote $P(t_k, t_l; s_k)$ the discount factor from time t_k to time t_l at scenario $s_k \in S_k$. For easier formulation, we divide payments between assets $A_{t_k}(s_k, e)$ and liabilities $L_{t_k}(s_k, e)$. These can be calculated as follows:

$$A_{t_k}(s_k, e) = \sum_{\substack{t_i, t_j : \\ t_i \le t_k < t_j}} P(t_k, t_j; s_k) w_{t_i,t_j}(t_j; a_i(s_k), e) + \mathbb{I}_{[t_k < t(e)]} \sum_{t : t_k < t \le T} P(t_k, t; s_k)\pi,$$

$$L_{t_k}(s_k, e) = \sum_{\substack{t, t_i, t_j : \\ t_i \le t_k < t \le t_j}} P(t_k, t; s_k) u_{t_i,t_j}(t; a_i(s_k), e) + \sum_{\substack{t_i, t_j : \\ t_i \le t_k < t_j}} P(t_k, t_j; s_k) v_{t_i,t_j}(t_j; a_i(s_k), e).$$

Bringing these formulas together with the amount of money left on current account $B_{t_k}(s_k, e)$ gives us value of the portfolio $V_{t_k}(s_k, e)$ at scenario (s_k, e). That reads as

$$V_{t_k}(s_k, e) = B_{t_k}(s_k, e) + A_{t_k}(s_k, e) - L_{t_k}(s_k, e).$$

This leads us to the formulation of the objective function $f(r, u, v, w)$, which expresses the value of the loan at the final time horizon. We have

$$f(r, u, v, w) = p(r) \cdot \sum_{s_K \in S_K, e \in E} p(s_K, e) V_{t_K}(s_K, e), \tag{11}$$

where variables u, v, w symbolically stand for the sets of decision variables as defined above.

To complete the model formulation, we need to specify the final set of constraints. It will consists from the already mentioned non-anticipativity constraints, as we need to make sure that the decisions of the company in times and scenarios, where the event has not been revealed yet are the same. We impose

$$u_{t_i,t_j}(s_i, e_1) = u_{t_i,t_j}(s_i, e_2) \quad v_{t_i,t_j}(s_i, e_1) = v_{t_i,t_j}(s_i, e_2) \quad w_{t_i,t_j}(s_i, e_1) = w_{t_i,t_j}(s_i, e_2),$$
$$\forall s_i \in S_i, e_1, e_2 \in E : t_i < \min\{t(e_1), t(e_2)\}, t_j > t_i, i = 0, \ldots, K.$$

This leads us to a formulation of a stochastic program

$$\max_{r,u,v,w} \quad f(r, u, v, w), \qquad \text{s.t. } \Big(u_{t_i,t_j}(s_i, e), v_{t_i,t_j}(s_i, e), w_{t_i,t_j}(s_i, e)\Big) \in \Theta(s_i, e), \quad i = 0, \ldots, K,$$

where $\Theta(s_i, e)$ represents the set of feasible decisions defined by the above-specified constraints for each (s_i, e).

3 Numerical Results

In this part, we will shortly summarise one set of results which we obtained. These will be most importantly used to stating other research questions which could be later investigated. We set the time horizon to be 5 years with decisions once a year and we enabled to the company to borrow and lend money for a period of one to five years. The branching was chosen to be $5 - 4 - 3 - 2 - 1$, leading to $|S_K| \cdot |E| = 120 \cdot 10 = 1200$ scenarios. Notional was set to 50000 CZK. The prepayment and default probabilities were set according to market data.

The optimal solution of the program was to offer to the client a rate $r = 7.056\%$, with $p(r) = 0.486$, while the expected value solution suggested the one of $r = 7.415\%$ with $p(r) = 0.398$ to be the most profitable. Expected value problem was formulated as an optimization problem where client cash-flows were weighted by their probabilities and the company replicated client's loan. The optimal borrowing and lending strategy of the company was basically to close only one year loans and if spare money were available, then the company should have lend them for the longest period possible (until the final time horizon). That strategy came as a no surprise, as it holds for a loan the shorter, the cheaper. The optimal value of the program was 1153 CZK, while expected value solution had objective value of 657 CZK.

To analyse results, we investigated loan performance in different interest rate scenarios for different events. First, we divided interest rate scenarios into three groups — *high, middle, low*, depending on their first-stage node

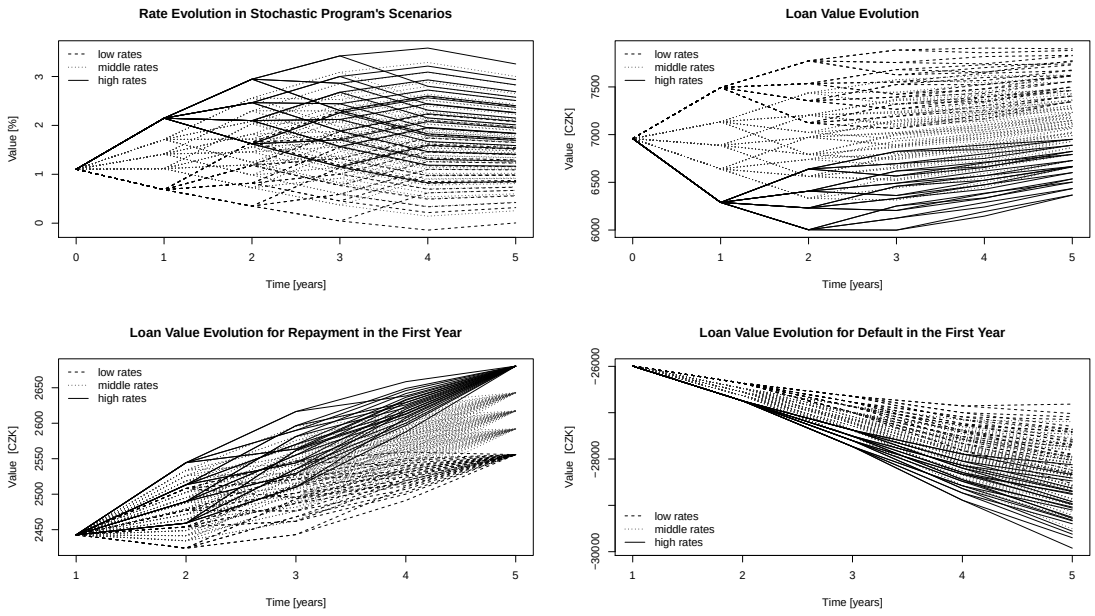


Figure 1 Interest rate and the optimal loan value evolution in time.

(see topleft figure in Figure 1). Then, we looked into performance when the client complied with the original terms of the loan. Because of the one-year borrowing strategy, the company profits on interest rate decrease and it loses money when it increases. However, the company uses *high* interest rate environment to compensate a bit for the initial loss by borrowing money earned from the loan and it earns high interest. Bottom figures in Figure 1 show performance of a loan when it is fully repaid/defaulted in the first year. There, one can see that high interest rate environment is preferred for loan repayment, because company then can reinvest money for higher interest. Unfortunately, as we mentioned earlier, these scenarios have lower probability (full repayments are more likely to occur when interest rate decreases). It is a complete opposite for the case of client defaults, as then the company is required to borrow additional money to finance its liabilities and that costs more in higher interest rate environment.

4 Conclusion

We formulated a non-linear multi-stage decision-dependent-randomness stochastic program describing a life cycle of a consumer loan. We found that the here-and-now solution to borrow money only for one year allows more flexibility when dealing with full prepayments and although the proposed optimal offered rate to the client was lower than the one of the expected value solution, the optimal value was almost twice as large. This underlines the usability of stochastic programming in finance, especially in asset-liability management.

We have also investigated the performance of the optimal solution in stochastic program's scenarios and we identified places which shall be addressed in a goal to improve interest rate risk and credit risk parameters of such a loan. Under a word *interest rate risk*, we mean to control the difference in loan performance between *high* and *low* rates, and also for some extreme interest rate movements, for the same events. On the other hand, we could imagine that *credit risk* would mean to impose some constraint on possible losses which would be caused by client default. These could be incorporated for example by a limit on capital requirement stemming from such a loan or by some penalty which would be imposed for having too large exposure to client default by the company. Implementation of various risk constraints or other modifications to control for risks faced during the life of a loan are required for efficient use of the program. Another natural extension which would be very useful is reformulating the program in order to comply with usual rules of mortgages; i.e. most importantly, to implement the refix property into the program. This extension appears very challenging to us as it would rapidly increase the complexity of the program.

Acknowledgements

The work was supported by the grants GAUK No. 305-10/250739 and GAČR No. 18-05631S. Their support is greatly acknowledged.

References

[1] D. R. Carino, T. Kent, D. H. Myers, C. Stacy, M. Sylvanus, A. L. Turner, K. Watanabe, and W. T. Ziemba. The Russell-Yasuda Kasai Model: An Asset/Liability Model for a Japanese Insurance Company Using Multistage Stochastic Programming. *Interfaces*, 24(1):29–49, 1994.

[2] M. A. H. Dempster and A. M. Ireland. MIDAS: An expert debt management advisory system. In *Data, expert knowledge and decisions*, pages 116–127. Springer, 1988.

[3] C.L. Dert. *Asset Liability Management for Pension Funds: A Multistage Chance Constrained Programming Approach*. PhD thesis, Erasmus University, Rotterdam, 1995.

[4] J. Dupačová and M. Kopa. Robustness in stochastic programs with risk constraints. *Annals of Operations Research*, 200(1):55–74, 2012.

[5] J Dupačová and M. Kopa. Robustness of optimal portfolios under risk and stochastic dominance constraints. *European Journal of Operational Research*, 234(2):434–441, 2014.

[6] A. Geyer and W. T. Ziemba. The Innovest Austrian pension fund financial planning model InnoALM. *Operations Research*, 56(4):797–810, 2008.

[7] K. Hoyland. *Asset Liability Management for a Life Insurance Company: A Stochastic Programming Approach.* PhD thesis, Norwegian University of Science and Technology, 1998.

[8] M. Kopa, V. Moriggia, and S. Vitali. Individual optimal pension allocation under stochastic dominance constraints. *Annals of Operations Research*, 260(1-2):255–291, 2018.

[9] M. I. Kusy and W.T. Ziemba. A Bank Asset and Liability Management Model. *Operations Research*, 34(3): 356–376, 1986.

[10] S. R. Pliska and J. Ye. Optimal life insurance purchase and consumption/investment under uncertain lifetime. *Journal of Banking & Finance*, 31(5):1307–1319, 2007.

[11] S. Vitali, V. Moriggia, and M. Kopa. Optimal pension fund composition for an italian private pension plan sponsor. *Computational Management Science*, 14(1):135–160, 2017.

Application of two distributions mixture to testing hypothesis in auditing

Grzegorz Sitek[1]

Abstract. The considered problem can be treated as a particular topic in the field of testing some substantive hypothesis in financial auditing. The main theme of the paper is the well-known problem of testing hypothesis on admissibility of the population total of accounting errors amounts. In auditing, some population elements contains no errors, whereas other population elements contains errors of varying amounts.
Usually, statistical inference considered in auditing is focused on estimation of the total or the mean error amount. The inference is based on data observed in samples drawn from a population of documents. An audit sample delivers two pieces of information: the book (recorded) amount and the audited (correct or true) amount. Let us note that a value of the book amount can be a sum of smaller book amounts. The book values will be treated as observation of a random variable distributed according to mixture of two probability distribution. The mixing coefficient is equal to the proportion of the items with non-zero errors amounts. The well-known method of moments is proposed to estimate mixtures of probability distribution. It let us construct some statistic to test the outlined hypothesis. We use the Monte Carlo method in estimating the power of a test.

Keywords: accounting error, mixture of probability distribution, Monte Carlo tests, statistical auditing.

JEL classification: 60E05
AMS classification: 62P99

1 Introduction

Financial accounting is concerned with the collection of data about the economic activities of a given firm and the suumarizing and raporting of them in the form a financial statements. Auditing, on the other hand, refers to the independent verification of the fairness of these financial statements. The auditor collects data that useful for verification from several sources and by different means. An item audit sample produces two pieces information, namely the book (recorded) amount and the audited (correct) amount. The difference between two is called the error amount. The foregoing studies provide some insight into the evidence on the characteristics of errors in auditing populations. Stringer[5] and Kaplan[4] shows, that the accounting populations are highly positively skewed, and there is considerable diversity in the characteristics of error amounts in accounting populations across the accounting subsystem. The studies suggest that there is considerable variation in each account type, error amount distribution as well as in error rates and balance of errors between overstatement and understatement. Moreover, it means that correlation coefficient between the sample mean and variance of book amounts is proportional to positive value of the third central moment of the book amounts distribution. We can see it in [1] and [3]. Hence, that is why distributions of accounting amounts, expenditures or income are modelled e.g. by means of gamma, Poisson or lognormal probability distributions. Below we consider a problem of testing appropriately formulated statistical hypotheses about admissibility of the total or the mean accounting errors.

2 Auditing error modeling using mixtures

In [7] Wywial proposed following model. Let U be the population of accounting documents of size N where accounting amounts are observed. Some of them are contaminated by errors. In the population U there are book amounts x_i observed for each population element $i \in U$. Let $x^T = [x_1, \ldots, x_N] \in R_+^N$ be the observation of the random vector $X^T = [X_1, \ldots, X_N]$. The true (without errors) accounting amounts are denoted by the values y_i, $i \in U$ and let $y^T = [y_1, \ldots, y_N] \in R_+^N$ be the value of the random vector $Y^T = [Y_1, \ldots, Y_N]$. The vector of accounting amounts contaminated by errors will be denoted by $w^T = [w_1, \ldots, w_N]$ and w is a value of $W^T = [W_1, \ldots, W_N]$. Finally, let $Z^T = [Z_1, \ldots, Z_N\ ]$, where $Z_i = 0$ ($Z_i = 1$) if $X_i = Y_i$ ($X_i \neq Y_i$), $i \in U$.

Let $F_0(y|\theta_0)$ be the probability distribution function of the random variable Y, whose values are true accounting and $\theta_0 \in \Theta_0$ where Θ_0 is the parameter space. The distribution function of W is denoted by $F_1(w|\theta_1)$, where

[1]University of Economics in Katowice, Faculty of Management, Department of Statistics, Econometrics and Mathematics, grzegorz.sitek@ue.katowice.pl

$\theta_1 \in \Theta_1$. Moreover, let $\Theta = \Theta_0 \cup \Theta_1$. We assume that an accouting errors apperrs with probability p. We can write $Z = 1$ when an accouting error occurs $P(Z = 1) = p$ and $Z = 0$ when it does not occur $P(Z = 0) = 1 - p$.

Let the values of the random variable X be the observation of acounting amounts generated as follows. The conditional distribution functions of X will be defined by: $F(X|Z = 0) = F_0(y|\theta_0)$ and $F(X|Z = 1) = F_1(w|\theta_1)$. According to the well-known total probability thorem we have: $F(x) = F(x \mid Z = 0) P(Z = 0) + F(x \mid Z = 1) P(Z = 1)$ and finally

$$F(x|\theta) = (1-p)F_0(x|\theta_0) + pF_1(x|\theta_1) \tag{1}$$

where $\theta = \theta_0 \cup \theta_1$ and $\theta \in \Theta = \Theta_0 \cup \Theta_1$ is the parameter space. Hance, the probability distribution of the observed accounting amounts is a mixture of the distribution function $F_0(x|\theta_0)$ of the true amounts and the distribution function $F_1(x|\theta_1)$ of the amounts contamineted by errors. When the random variables Y and W are continuous, by differentiating both sides of equation (1) we have

$$f(x|\theta) = (1-p)f_0(x|\theta_0) + pf_1(x|\theta_1) \tag{2}$$

Therefore, the probability density of the observed accouting amounts is a mixture of density $f_0(x|\theta_0)$ of the true amounts and density $f_1(x|\theta_1)$ of the amounts contaminated by errors. Let R and Y be independent where R is the accounting error. Hence $W = Y + R$, $X = Y + ZR$, $X = (1 - Z)Y + ZW$.
The basic moments of the random variable X are:

$$E(X) = (1-p)E(X|Z=0) + pE(X|Z=0) = (1-p)E(Y) + pE(W) \tag{3}$$

$$V(X) = p(1-p)(E(W) - E(Y))^2 + pV(W) + (1-p)V(Y) \tag{4}$$

In practice, all values of X are known before auditing process. Observations x of X are treated as a specific auxiliary data. Auditing process leads to observation of values Z_i, Y_i and W_i, $i \in U$. Let $\overline{X} = \frac{1}{N}\sum_{i\in U} X_i$, $\overline{Y} = \frac{1}{N}\sum_{i\in U} Y_i$, $\overline{W} = \frac{1}{N}\sum_{i\in U} W_i$. Their values will be denoted $\overline{x}$, $\overline{y}$, $\overline{w}$. Let an auditor arbitrary selects the sample s of the size n from U. Hence, X_s is the subvector of X, $n \leq N$. The random vector X_s is observed in s where the objects are controlled. After the auditing process the sample s is split into two disjoint sub-samples s_0 and s_1 where $s = s_0 \cup s_1$. The set s_1 is of size $n_1 = k$ and the set s_0 is of size $n_0 = n - k$. In the sub-sample s_0 here are observed accounting amounts without errors. They are values of the random variables denoted by $\{X_i = Y_i, i \in s_0\}$. In the subsample s_1 accounting amounts contaminated by errors are observed as values of the random variables $\{W_i, i \in s_1\}$. Hence, before auditing process we have observations of the following data:

$$X = (X_i : i \in U) = (X_s,\ X_{U-s})$$

where

$$X_s = (X_i : i \in s), \qquad X_{U-s} = (X_i : i \in U - s)$$

After the auditing process we have observations of the following data:

$$\aleph_U = (\aleph_s,\ X_{U-s}),\ \ \aleph_s = (X_i,\ Z_i) : i \in s) = (Y_{s_0},\ W_{s_1})$$

In the context of model approach our purpose is to test the hypothesis about the expected value of the following difference of the sum of observed in the population accounting amounts and the sum of the true values. On the basis of the equation (3) we have:

$$\tau = E\left(\overline{X} - \overline{Y}\right) = E(X) - E(Y) = p(E(W) - E(Y))$$

or

$$\tau(\theta) = E(X|\theta) - E(Y|\theta_0) = p(E(W|\theta_1) - E(Y|\theta_0)) \tag{5}$$

In [7] and [6] Wywiał described model of two Poisson distributions and mixture of gamma probability distributions respectively. Below the mixture of gamma and sum gamma and normal probability distributions is taken into account. Let $Y \sim G(a, c)$ and $R \sim N(\mu, \sigma)$ be independent and $W = Y + R$.

$$f(x|a, c, \mu, \sigma) = (1-p)f_0(x|a, c) + pf_1(x|a, c, \mu, \sigma) \tag{6}$$

where $f_1(x|a, c, \mu, \sigma)$ is the density of the variable W and

$$f_0(x \mid a, c) = \frac{c^a}{\Gamma(a)} x^{a-1} e^{-cx}\ x > 0 \tag{7}$$

Based on expressions (3)-(4) we obtain:

$$E(X) = \frac{a}{c} + p\mu\ V(X) = p(1-p)\mu^2 + p\sigma^2 + \frac{a}{c^2} \tag{8}$$

In the next sections parameters of the distributions mixture will be estimated by means of the well-known method moments.

3 Hypothesis in auditing

Let $\tau = E(\overline{X} - \overline{Y})$ be the expected mean accounting error. Audit purpose is inference on τ or on the expected total accounting error $N\tau = E(\sum_{i\in U} X_i - \sum_{i\in U} Y_i)$. In particular, when we assume that τ_0 is the admissible mean accounting error then the inference reduces to testing the following hypothesis:

$$H_0 \ : \ \tau \leq \tau_0 \, H_1 \ : \ \tau > \tau_0 \tag{9}$$

Below the probability distribution of X_i is treated as the mixture of distributions of X_i and Y_i, $i \in U$.

3.1 Inference based on sample moments

Now let us consider the hypotheses denoted by the expression (9) about the function of the parameters $\tau(\theta) = \tau(p, a, c, \mu, \sigma)$. In our case on the basis of the expressions (5) and (8) we have:

$$\tau(\mu, p) = E(X|a, c, \mu, \sigma, p) - E(Y|a, c, p) = p\mu \tag{10}$$

We assume that $n-k > 0,\ k > 0$ and $N-n > 0$. In this situation after the auditing process we have observations of data $\aleph_U$ where $s \neq \emptyset$ and $s_1 \neq \emptyset$. In this case it is possible to simplify the estimation of the parameter in the following way:

$$\begin{cases} E\,(X) = \frac{a}{c} + p\mu \\ E\,(Y) = \frac{a}{c} \\ E\,(R) = \mu \\ V\,(X) = \frac{a}{c^2} \\ V\,(R) = \sigma^2 \end{cases} \tag{11}$$

Let $\overline{y}_{s_i} = \frac{1}{n_i}\sum_{k\in s_i} y_k \ v_{s_i} = \frac{1}{n_i}\sum_{k\in s_i}(y_k - y_{s_i})^2$. After replacing the moments $E(X)$, $V(X)$, $E(Y)$, $E(R)$, and $V(Y)$ with the sample moments $\overline{X}_{U-s}$, $V_{U-s}(X)$, $\overline{Y}_{s_0}$, $\overline{R}_{s_1} = \overline{W}_{s_1} - \overline{Y}_{s_1}$ and $V_{s_0}(Y)$, respectively, the appropriate algebraic operations let us evaluate the following estimators of parameters p, a, μ, c and σ.

$$\begin{cases} P_U = \frac{\overline{X}_{U-s} - \overline{Y}_{s_0}}{\overline{R}_{s_1}} \\ A_{s_0} = \frac{\overline{Y}^2_{s_0}}{V_{s_0}(Y)} \\ \mu_{s_1} = \overline{R}_{s_1} \\ C_{s_0} = \frac{\overline{Y}_{s_0}}{V_{s_0}(Y)} \\ \sigma^2 = \overline{R^2}_{s_1} - \overline{R}^2_{s_1} \end{cases} \tag{12}$$

provided denominators of the above fractions are positive. In this case the unbiased estimator of τ is as follows:

$$\widehat{\tau}_2 = \overline{X}_U - \overline{Y}_{s_0} \tag{13}$$

This let us construct the following test statistic of hypothesis H_0

$$\hat{G}_2 = \frac{\widehat{\tau}_2 - \tau_0}{\sqrt{\frac{V_U(x)}{N-n} + \frac{V_{s_o}(Y)}{n_0}}} \tag{14}$$

where $n_0 = n - k$. On the basis of the well known properties the probability distribution of sample moments we can prove that $\hat{G}_2 \sim N(0, 1)$ when H_0 is true, $N \to \infty,\ n_0 \to \infty$. We can see it in [1].

3.2 Monte Carlo tests

In case of moderate and small sample sizes the hypotheses can be tested by means of Monte Carlo procedure. We can see Monte Carlo procedure in [2]. In order to apply this method all of the considered models should be reparametrised. This let us simplify the inference on total error amount.

Our purpose is testing hypothesis $H_0 : \tau = \tau_0$ against alternative hypothesis $H_1 : \tau > \tau_0$. Parameter a, c, σ, p and τ are estimated by means of statistics defined by expressions (12) and (13).

- On the basis of the equation $\tau = p\mu$ we replace the parameter $\mu = \frac{\tau}{p}$
- We transform the model, given by (6), to the following:

$$f(x|a, c, \tau, \sigma, p) = \ (1-p)f_0(x|a, c) + \ pf_1(x|a, c, \tau, \sigma)$$

- We generate data $\aleph^{0,i} = \left(Y_{s_0}^{0,i},\ W_{s_1}^{0,i},\ X_{U-s}^{0,i}\right)\ for\ i = 1,\ldots,m$ according to densities $f_0(y|a_{s_0}, c_{s_0})$, $f_1(w|a_{s_0}, c_{s_0}, \tau_0, \sigma_s)$ and $f(x|a_{s_0}, c_{s_0}, \tau_0, \sigma_s, p_U)$ and data $\aleph^{1,i} = \left(Y_{s_0}^{1,i},\ W_{s_1}^{1,i},\ X_{U-s}^{1,i}\right)\ for\ i = 1,\ldots,m$ are generated based on $f_0(y|a_{s_0}, c_{s_0})$, $f_1(w|a_{s_0}, c_{s_0}, \tau_1, \sigma_s)$ and $f(x|a_{s_0}, c_{s_0}, \tau_1, \sigma_s, p_U)$
- We evaluate the following: $\hat{g}_2^{(e,i)} = \frac{\tau^{(e,i)} - \tau_0}{\sqrt{\frac{v_U(X^{(e,i)})}{N-n} + \frac{v_{s_o}(Y^{(e,i)})}{n_0}}}\ \tau^{(e,i)} = \overline{X}_{U-s}^{(e,i)} - \overline{Y}_{s_0}^{(e,i)}$ e=0,1
- The sequence $\left\{\hat{g}_2^{(e,i)},\ i = 1,\ldots,m\right\}$ approximates the distribution of the test statistic $\hat{G}_2$ under the assumption that H_e is true, $e = 0, 1$.
- p-value of the test denoted by Dufour and Khalaf in [2] by $\widehat{\alpha}$

$$\widehat{\alpha} = \eta_e,\ e = 0,\ \eta_e = \frac{mw_e}{m+1}\ ,\ w_e = \frac{1}{m}\sum_{i=1}^{m} I\left(\hat{g}_2^{(e,i)}\right),$$

$$I\left(\hat{g}_2^{(e,i)}\right) = \begin{cases} 1\ if\ g > \hat{g}_2 \\ 0\ if\ g < \hat{g}_2 \end{cases}$$

- When $\widehat{\alpha} < \alpha$ hypothesis H_0 is rejected but this decision can be wrong with the risk of incorrect rejection equal to probability α
- When $\widehat{\alpha} > \alpha$ hypothesis H_0 is accepted and this decision can be wrong with the risk of incorrect acceptance equal to $\left(1 - \widehat{\beta}\right)$

Example 1. The population of $N = 4000$ invoices from an anonymous firm is considered. The simple sample of $n = 200$ invoices was selected in order to estimate parameters of the model. After auditing $n_1 = 62$ and $n_0 = 138$. The hypotheses are: $H_0 : \tau = \tau_0$ against $H_1 : \tau = \tau_1 = k\tau_0,\ k > 1$.

α/k	k=1.1	k=1.2	k=1.3
0.05	0.062	0.087	0.111
0.1	0.129	0.166	0.221
0.2	0.232	0.284	0.347

Table 1 Powers of the test n=200

The hypotheses are: $H_0 : \tau = 800$ against $H_1 : \tau = 960 = 1.2 \cdot 800$. Value of the test statistic, defned by (14) is: $\hat{g_2} = 0.0008$. Next, according to the simulation procedure the data were replicated 1000 times. This, under the assessed significance levels 0.2 let us evaluate the critical values equal to 0.902. Hence, under these significance levels hypothesis H_0 is not rejected. For the considered sequence of significance levels the approximated powers of the test is: 0.284. Therefore, for instance even for $\alpha = 0.2$ the acceptance of H_0 could be wrong decision with probability 0.716. Hence, size of the sample should be increased.

Example 2. The population of $N = 4000$ invoices from an anonymous firm is considered. The simple sample of $n = 400$ invoices was selected in order to estimate parameters of the model. After auditing $n_1 = 124$ and $n_0 = 276$. The hypotheses are: $H_0 : \tau = \tau_0$ against $H_1 : \tau = \tau_1 = k\tau_0,\ k > 1$.

α/k	k=1.1	k=1.2	k=1.3
0.05	0.081	0.118	0.181
0.1	0.146	0.231	0.318
0.2	0.287	0.379	0.493

Table 2 Powers of the test n=400

The hypotheses are: $H_0 : \tau = 800$ against $H_1 : \tau = 960 = 1.2 \cdot 800$. Value of the test statistic, defned by (14) is: $\hat{g_2} = 0.002$. Next, according to the simulation procedure the data were replicated 1000 times. This, under the assessed significance levels 0.2 let us evaluate the critical values equal to 0.904. Hence, under these significance levels hypothesis H_0 is not rejected. For the considered sequence of significance levels the approximated powers of the test are: 0.379. Therefore, for instance even for $\alpha = 0.2$ the acceptance of H_0 could be wrong decision with probability 0.621. In the case of a larger sample, we obtained a higher test power, however the probabilities of incorrect acceptance H_0 hypothesis are still too high.

4 Conclusions

Inference on the total error amount is based on appropriately evaluated confidence intervals. Of course they are related closely to testing problems but not exactly. In this paper the decision-making process in auditing is treated as a problem of testing appropriately formulated statistical hypotheses about admissibility of the total or the mean accounting errors. This approach let us control not only significance level (risk of incorrect rejection) but also probability of appearing the type II error (risk of incorrect acceptance) during testing the hypotheses. In this paper it was shown how to estimate parameters of interest by means of the method of moments. We use the Monte Carlo method in estimating the power of a test.

References

[1] Cramér H. (1962). *Mathematical Methods of Statistics*. Asia Publishing House, Bombay Calcutta New Delhi Madras London New York.

[2] Dufour J. M., Khalaf L. (2001). Monte Carlo test methods in econometrics. *Companion to Theoretical Econometrics*, ed. B. Baltagi. Oxford, U.K., 494–519.

[3] Frost P. A., Tamura H. (1986). Accuracy of auxiliary information interval estimation in statistical auditing. *Journal of Accounting Research*, 24, 57–75.

[4] Kaplan, R.S. (1973). Statistical Sampling in Auditing with Auxiliary Information Estimators. *Journal of Accounting Research*, 238–258.

[5] Stringer, K.W. (1963). Practical Aspects of Statistical Auditing. *Preceeding of Business and Economic Statistics Section of the American Statistical Association*, 405–411.

[6] Wywiał, J. L. (2018). Application of two gamma distributions mixture to financial auditing. *Sankhyã B: The Indian Journal of Statistics.*

[7] Wywiał, J. L. (2016). *Contributions to Testing Statistical Hypotheses in Auditing*. PWN, Warsaw, 91-95.

Input-output analysis of Czech export

Jaroslav Sixta[1], Jakub Fischer[2]

Abstract. Input-Output tables provide specific tools for deep economic analyses. A fundamental question is how exports contribute to the economic growth. Standard approach based on the analysis of the export share on gross domestic product became not valid in the modern globalised world. This is more visible for the small and opened countries such as the Czech Republic. The approach developed by the OECD allows clear identification of the value added contained in exports. This approach can be extended to the level of the regions with respect to the developed regional input-output tables. The paper provides an analysis of the contribution of exports to the economic growth in the Czech Republic and its regions. Used methods are based on input-output methodology with some specific extensions.

Keywords: Input-Output, Exports, Regional.

JEL Classification: C67, O11
AMS Classification: 65C2

1 Introduction

Czech export belongs to the most discussed issue in our economy. The debates ranging from exports to gross domestic product ratio to the product structure and territorial composition. In a globalized world, exports are very linked to imports since lots of intermediates are required by globalized producers. It is useful to estimate composition of Czech exports in relation to imports. How much imports are required for exports? Among different possibilities, we prefer input-output analysis since it is a suitable tool for these computations. In a simple extension, the input-output analysis allows to estimate both direct and indirect inputs.

Similar analyses are being conducted on the national level but they can be easily enlarged to the regional level. When regional input-output tables are available, the analysis can be performed on both single regions and multiple regions matrices. The modern use of such phenomena relates mainly to the work of the OECD – Trade in Value Added (TiVA), [5] that comprise huge database of countries' data. The data offered by the OECD comprise countries and similarly that can be extended to the regions of a specific country. In our case for the Czech Republic.

Obvious dependency of the Czech Republic on exports can be distributed into the regions to asses which are the most involved regions. The basic data source is formed by the regional input-output tables (RIOTs) that are arranged into the form of inter-regional model. These 14 input-output tables come from the research of the University of Economics and can be downloaded for free, see kest.vse.cz. These academic based RIOTs were designed to fit official symmetric input-output tables (SIOTs) regularly compiled by the Czech Statistical Office and can be regarded as an extension of available data source for regional analyses. RIOTs represent often academic exercise, see Többen and Konenberg [8]. The latest RIOTs for the Czech Republic are available for 2013 even though national SIOTs currently exists for 2015, as well. The valuation is standard, basic prices and the dimension of 82 products is more than sufficient. The 14 RIOTs correspond to the regions of the Czech Republic on the level of NUTS 3. The tables were arranged into a form of inter-regional model by the Newton gravity approach, see Sixta and Šafr [6]

2 Methodology and data

The data comes from two sources, the first is the Czech Statistical Office with official figures for exports, SIOTs and national accounts, see [2]. The second source is based on the academic research of the University of Economics that offers RIOTs and the inter-regional input-output model derived according to the Newton gravity approach [3].

The main approach to our analysis consists in the estimations of the share of gross value added in exports. It refers to value created with the country or regions. As other computations it refers to standard input-output analysis, see Miller and Blair [4]. It means that lots of exports contain direct or indirect imports in terms of interme-

[1] University of Economics, Department of Economic Statistics, W. Churchilla 4, Prague, sixta@vse.cz
[2] University of Economics, Department of Economic Statistics, W. Churchilla 4, Prague, fischerj@vse.cz

diates and this is a possible way, how to avoid it. The computation is based on the input-output multipliers. Firstly, we computed multipliers of exports as

$$m_j^v = \sum_{i=1}^{n} l_{i,j}^m \frac{v_j}{x_j}, \tag{1}$$

where

m.. Leontief multiplier of exports and upper index *v* stands for label – value added,

l.. elements of Leontief invers matrix,

v.. gross value added,

x.. output.

For computational purposes it can be rewritten into matrix form as:

$$\mathbf{m}^v = (\mathbf{h}^v)^{\mathbf{T}} \mathbf{L},$$

$$h_j^v = \frac{v_j}{x_j}, \tag{2}$$

where

m.. vector of Leontief multiplier of exports,

L.. Leontief invers matrix,

h.. auxiliary vector.

The usage of the multipliers can be interpreted as the change of the indicator according to a given incentive. This is also used for the estimation of the contained gross value added in exports on the both regional and inter-regional level. The difference lies in the fact that inter-regional model allows the interaction between all regions. For the nominal estimates, the multipliers must be multiplied by exports.

3 Dependency on exports

The Czech Republic is crucially dependent on exports, it is easy to prove it since the exports to gross domestic product ratio has been rising since 2000. The inflow of foreign direct investment led to the higher connection of resident companies to foreign trade. In 1990s the share was about 40% and in 2016 it is more that 80%. Such dependency is redeemed by both the increase of imports and outflow of primary incomes (dividends and reinvested earnings), see figure 1.

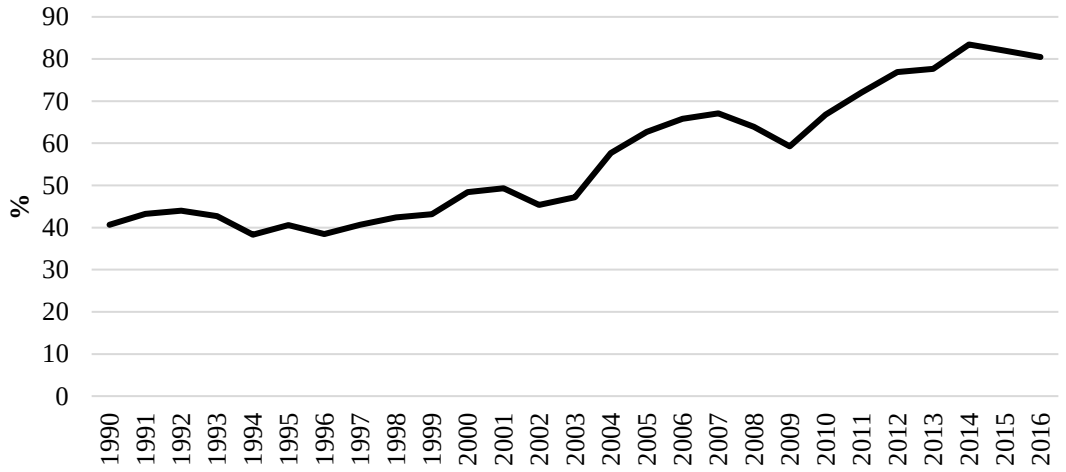


Figure 1 Exports as a % of GDP

The structure of Czech exports plays important role since the highest share belongs to manufacturing industry. Despite minor changes in the structure, the Czech economy became very dependent on the car industry. Simple structures derived from Czech supply and use tables (figure 2) illustrate that exports of the cars and related products increased nearly fivefold between 2000 and 2016. On the contrary, exports of computers and similar IT products remained steady between 2010 and 2016. We can find lots of other products being exported from the Czech Republic, but cars are dominating. In 2006, exports of cars and related products took 24% of total Czech exports but only 15% in 2000.

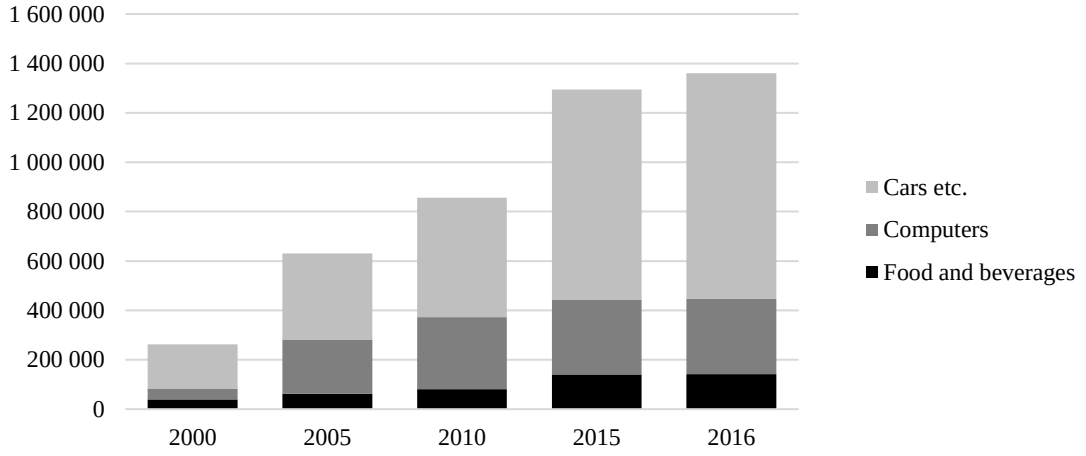


Figure 2 Selected exports product, current prices

4 Regional contribution to exports

4.1 Simple analysis of regional export contribution

The dependency of the Czech Republic on exports can be easily illustrated but the different issue is the impact on the Czech regions. Exports can be directly linked to the regions of the origin. Another issue is the indirect impact (section 4.2). The contribution of Czech regions to exports is estimated rather simple, as a share of exports excluding inter-regional flows on total exports of the Czech Republic, see table 1. The data comes from regional input-output table, see Sixta and Vltavská [7].

Region	Label	2011	2013
Hlavní město Praha	PHA	13.0%	11.3%
Středočeský kraj	STC	16.3%	16.7%
Jihočeský kraj	JHC	4.4%	4.5%
Plzeňský kraj	PLZ	5.3%	6.1%
Karlovarský kraj	KAR	2.0%	1.3%
Ústecký kraj	UST	7.1%	6.9%
Liberecký kraj	LIB	3.8%	3.9%
Královéhradecký kraj	KRH	5.3%	5.0%
Pardubický kraj	PAR	7.1%	6.9%
Kraj Vysočina	VYS	4.2%	4.2%
Jihomoravský kraj	JHM	8.6%	10.8%
Olomoucký kraj	OLM	4.1%	3.9%
Zlínský kraj	ZLN	5.5%	6.2%
Moravskoslezský kraj	MRS	13.4%	12.6%
Czech Republic	Total	100.0%	100.0%

Table 1 Regional contribution to exports

The dominant position of Středočeský region increased from 16.3% to 16.7%. Besides we can find at least two other regions contributing significantly to exports (Prague and Moravskoslezský region) but their shares decreased. The highest increase is observed for Jihomoravský region where its share rose from 8.6% to 10.8%.

The structure of products exported significantly differs among the regions. The most involved region, Středočeský region is mainly connected with car industry and car industry is also connected with Moravskoslezský region. For example Pardubický region is close connected with exports of the IT products. The capital Prague is connected mainly with exports of services ranging from trade to other business services. As seen below, the crucial information does not lie in the exports but in the value added embodied in exports.

4.2 Input-Output analysis of regional exports

More detailed approach offers input-output analysis of embodied value added in exports. In line with formula 2, we can compute export multipliers and when these multipliers are multiplied by the exports, the value added content of exports is observed. It means that dependency of the Czech Republic on exports can be easily illustrated by simple indicators but the different issue is the impact on the value added of both the country and the region. The estimates of valued added embodied in exports expressed as a share of regional gross value added are presented in Table 2. The figures are influenced by the size of regional value added and hence Prague and Středočeský region are not on the top.

Region	Label	
Hlavní město Praha	PHA	29.4%
Středočeský kraj	STC	31.3%
Jihočeský kraj	JHC	36.0%
Plzeňský kraj	PLZ	37.0%
Karlovarský kraj	KAR	39.0%
Ústecký kraj	UST	40.4%
Liberecký kraj	LIB	42.0%
Královéhradecký kraj	KRH	43.1%
Pardubický kraj	PAR	44.5%
Kraj Vysočina	VYS	45.9%
Jihomoravský kraj	JHM	46.0%
Olomoucký kraj	OLM	46.3%
Zlínský kraj	ZLN	46.8%
Moravskoslezský kraj	MRS	47.3%
Czech Republic	Total	29.4%

Table 2 The share of value added embodied in exports on regional gross value added.

On the contrary, if we describe value added embodied in exports in relation to exports, different explanation is provided, see Figure 3. It shows that higher value added connected with exports can be found in the capital Praha (PHA) and Středočeský region (STC). On the case of Středočeský region, we can easily describe the difference between exports and gross value added embodied in exports. The difference is very significant also for Moravskoslezský, Jihočeský and Ústecký regions. The estimates of gross value added embodied in exports incorporate also the links between all other regions. One of the highest differences can be found for Pardubický region where lots of IT products origin. The details about ICT products and services are deeply elaborated in Fischer and Vltavská [1]. In many cases, it is rather assembling from imported parts than knowledge-based IT production. The reasons for these differences can be different ranging high demand for imports in the supply chain of multinational enterprise to very simple assembling imported products.

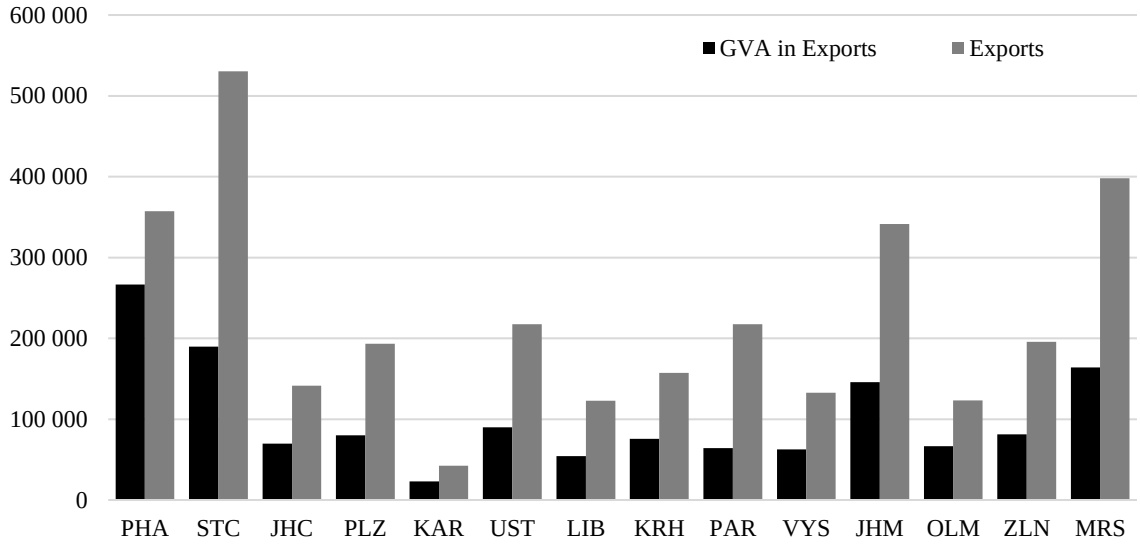


Figure 3 Gross value added embodied in exports and regional exports

5 Conclusion

The input-output tools allow deep and detailed analysis of the economy. It is obvious that the data availability is a necessary precondition. The input-output tables belong to the standard set of economic information provided by national accounts, but its regional extension usually belongs to the area of research. That is also the situation in the Czech Republic. Regional input-output tables come from the research conducted at the University of Economics in Prague, but they are very closely linked to official national data for the Czech Republic.

The analysis of the Czech Economy from the perspective of exports' contribution to gross value added brings additional information for exports strategies. It sounds logical that countries such as the Czech Republic should promote companies in the creation of business bring significant value added. The time when Central and East European countries supported multinational enterprises in creating low-wage jobs such as pure assembling without any research has passed. The difference between exports and value added embodied in exports is significant in many cases. There are two issues that should be simultaneously tackled the first is connected with production, exports and value added and the second is connected with the heterogeneity of exports. The diversification of the production could potentially limit economic problems in the future. The emphasis of the Czech economy on the car industry may be dangerous once.

The input-output tools and regional input-output tables provide powerful analytical options. Regional input-output tables can be downloaded for free from kest.vse.cz. These tables for 2013 are being updated to fit the latest analytical necessities.

Acknowledgements

Supported by the Institutional Support for Long Period and Conceptual Development of Research and Science at Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Fischer, J.; Vltavská, K. *National accounts as a useful data source for competitiveness analysis including the ICT impact*. Jindřichův Hradec 07. 09. 2011 – 09. 09. 2011. In IDIMT–2011. Linz: Trauner Verlag universitat, 2011, s. 141–146. ISBN 978-3-85499-873-0.

[2] Kahoun, J., Sixta, J.: Regional GDP Compilation: Production, Income and Expenditure Approach. Sta*tistika* (2013), 4, 24–36.

[3] Gawthorpe, K., Šafr, K. *Regional & Multi-Sector Input-Output Model for the Czech Republic*. In: 20th International Scientific Conference AMSE. Szklarska Poreba, 30.08.2017 – 03.09.2017. Szklarska Poreba: Publishing House of Wroclaw University of Economics, 2017, s. 169–182. ISBN 978-83-7695-693-0.

[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, Šafr, K. *Regional Input-Output Tables and Inter-Regional Model for the Czech Republic*. In: Input-Output-Workshop Spezial 2017. Osnabrück, 30.03.2017 – 31.03.2017. Osnabrück: GWS, 2017. 4 s.
[7] Sixta, J., Vltavská, K. Regional Input-output Tables: Practical Aspects of its Compilation for the Regions of the Czech Republic. *Ekonomický časopis*. (2016), roč. 64, č. 1, s. 56–69. ISSN 0013-3035.
[8] Többen J., Konenberg. T.H.: Construction of Multi-Regional Input-Output Tables Using the Charm Method. *Economic Systems Research* (2015), 4, 487-507.

Exchange Offices Cash Delivery Planning: A Vehicle Routing Problem Application

Veronika Skocdopolova[1], Denisa Mouchova[2]

Abstract. Exchange offices sell and buy notes of different currencies. To run their business they need to have enough cash every day. Distribution of cash has to respect some specific rules due to its safety. This paper deals with problem of cash delivery to 49 exchange offices in the Czech Republic using own cars and package sending by external company. The capacity of the available cars is large enough for the quantity of money that has to be distributed per day. Therefore it is not necessary to take the car capacity into consideration.
The problem presented in this paper is solved using modified vehicle routing problem with time windows. Vehicle routing problem is a widespread model used for planning routes for goods distribution. It is an NP-hard problem. The modification used in this paper lies in an extension, that the money can be distributed not only by own cars of the exchange company, but it can be send by special type of package. The aim of the model is to decide, which offices should be supplied by cars and which by sending the special package, and plan the car routes with minimal total costs of the distribution.

Keywords: Vehicle routing problem, time windows, integer programming, Gurobi.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

This paper deals with an application of a modified vehicle routing problem with time windows. The model is applied to cash distribution to exchange offices for a company in the Czech Republic.

1.1 Exchange Offices

The company we obtained the data from is a family company with solely Czech capital. It has been operating since 1993. The company provides various services. As an exchange office it offers transactions with 40 different currencies, such as traditional Euro, American Dollar or British Pound or exotic currencies like Thai Baht, Israeli Shekel or Emirate Dirham. Beside the exchange services the company provides also cashless exchanges, payments abroad via bank accounts or via the WESTERN UNION service.

At the exchange offices all over the Czech Republic the customers can buy the highway signs for European Countries (Czechia, Slovakia, Slovenia, Switzerland, Austria, Hungary), parking permissions for Viennese and Prague airport and also investment silver and gold. The company runs 49 branch offices in Czechia.

1.2 Vehicle Routing Problem

The Vehicle Routing Problem (VRP) is a NP-hard problem [1]. It is a modification of Travelling Salesman Problem (TSP). The objective of the TSP is to find the shortest cycle 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 cycle is called Hamiltonian cycle. 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 [4], [5]. 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 [2].

[1] University of Economics, Prague, Faculty of Informatics and Statistics, department of econometrics, W. Churchill sq. 4, Prague 3, Czech Republic, veronika.skocdopolova@vse.cz.
[2] University of Economics, Prague, Faculty of Informatics and Statistics, department of econometrics, W. Churchill sq. 4, Prague 3, Czech Republic, xmoud01@vse.cz.

2 Problem Description and Model Formulation

The problem this paper deals with is to plan routes of company vehicles delivering cash to 49 exchange offices all over the Czech Republic. Beside delivering the cash by cars there is a possibility to deliver the money via a package sent through an external company.

The company owns 4 passenger cars which can be used for cash distribution. These cars are at the garage at the company headquarters. This is the starting point of all the delivery routes. To deliver the cash with a car, the company needs two drivers and one member of a security service for each car. Each car is able to transport all the cash that has to be delivered to all the offices in one day. Therefore we do not have to take the car capacity into consideration. The advantage of using the company's car for the distribution is the possibility of taking the spare amounts of currencies from the offices back to the headquarters or to transport them to another office on the route. On the other hand, the main disadvantage of this type of transport are high costs and time-consuming routes to distant offices.

The other possibility of cash distribution is using an external service of package sending. This external company provides sending money to any of the offices for a uniform price. The main benefit of this way of distribution is that the price for sending a package is the same, when they send it to the office in a neighboring town as well as to the office at the other end of the republic, where the costs for delivering by car would be several times higher. The drawback is less flexibility of the packages. The external company will not take the spare currencies back to the headquarters.

Nowadays, the company plans the distribution without any sophisticated tool. There are two experienced employees who plan the routes every day. Due to a safety issue, the routes have to be changed regularly. This point of the problem is not taken into consideration in our model.

The main goal of the problem the company would like to analyze is to minimize the total costs of the cash distribution. The result of the analysis has to consist of the decision which offices should be supplied via external service and which should be supplied by own cars of the company. Also the optimal distribution routes for the company's car has to be planned. The cash has to be delivered to the offices within its opening hours.

For solving this problem, we formulated the following mathematical model. It is a modification of the standard vehicle routing problem with time windows. The model operates with following variables and parameters:

- x_{ij}^k – binary variable that is equal 1 when the vehicle k goes from the office i to office j, and 0 otherwise; $i, j = 1, 2, \ldots, n$, where n is the number of exchange offices plus the starting point;
- y_i – binary variable that is equal 1, when the cash to the office i is delivered by a package, and 0 otherwise;
- t_i^k – arrival time of the vehicle k to the office i;
- C_{ij} – cost of the transfer from the office i to office j;
- PP – price of sending a package to any of the offices;
- S_i – time needful for the operation at the office i;
- D_{ij} – time the vehicle spends on the way between the offices i and j;
- M – large number;
- A_i – beginning of the time window, when the vehicle has to arrive to the office i;
- B_i – end of the time window, when the vehicle has to finish servicing of the office i and leave it.

The objective of the model is to minimize the total costs of delivering the cash to branch offices:

$$z = \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} x_{ij}^k + PP \sum_{i=1}^{n} y_i \tag{1}$$

subject to several constraints.
All the branch offices ($i, j = 2, 3, \ldots, n$) has to be supplied with cash exactly once (via the car or a package):

$$\sum_{k=1}^{K} \sum_{j=1}^{n} x_{ij}^k + y_i = 1 \qquad i = 2, 3, \ldots, n, \tag{2}$$

$$\sum_{k=1}^{K} \sum_{i=1}^{n} x_{ij}^k + y_j = 1 \qquad j = 2, 3, \ldots, n. \tag{3}$$

Each vehicle *k* can leave the garage at most once:

$$\sum_{j=2}^{n} x_{1j}^{k} \leq 1 \qquad \forall k. \tag{4}$$

When the vehicle *k* arrives to office *j*, it has also to departure from the office:

$$\sum_{i=1}^{n} x_{ij}^{k} = \sum_{i=1}^{n} x_{ji}^{k} \qquad j = 2, 3, \dots, n, \forall k. \tag{5}$$

The cash has to be delivered to the offices within its opening hours. The opening hours are defined as time windows. The following set of constraints (6) is a modification of so-called sub-tour breaking constraints, where the variable t_i^k represents the arrival time of the vehicle *k* to the office *i*. If the vehicle *k* then continues to the office *j*, the arrival time to that office has to be higher (or at least equal) than the arrival time to office *i* increased by the time need for operation at the office *i* and time need for the way between the offices *i* and *j*:

$$t_i^k + S_i + D_{ij} - M(1 - x_{ij}^k) \leq t_j^k \qquad \forall i, \forall j, \forall k. \tag{6}$$

The vehicle *k* has to arrive to each office within its opening hours (the time window $\langle A_i; B_i \rangle$):

$$A_i \sum_{j=1}^{n} x_{ij}^{k} \leq t_i^k \leq (B_i - S_i) \sum_{j=1}^{n} x_{ij}^{k} \qquad i = 2, 3, \dots, n, \forall k. \tag{7}$$

Each vehicle starts its route at the time point 0:

$$t_1^k = 0 \qquad \forall k. \tag{8}$$

Decision variables x_{ij}^k and y_i are binary variables:

$$x_{ij}^k \in \{0, 1\} \qquad \forall i, j, k, \tag{9}$$

$$y_i \in \{0, 1\} \qquad \forall i. \tag{10}$$

3 Conclusions

The presented model was used for optimization of cash distribution with data provided by the exchange company. The data set consists of costs of the transfer between each pair of offices, times needed for transfer between each pair of offices, estimated times needed time for the operation at each office, and opening hours of each office. We have calculated with 49 offices and with the garage as a starting point of the car routes, and with 4 company's cars. The mathematical model was written in MPL user interface and solved with the Gurobi solver. As was mentioned in the introduction part, vehicle routing problem is an NP-hard problem. Therefore we expected problems with finding an optimal solution for the problem.

We stopped solving the problem after 16 hours without obtaining the optimal solution. We have obtained only a feasible solution, where the gap between the best solution found and the best bound was approx. 20 %. Nevertheless also this feasible solution give as satisfactory results. The total expanses of the routes planned by the model were approx. 10 % lower than the real costs. The package was send to 5 offices from 49. The other offices were supplied by three of the company's cars. A conclusion of this result can be, that the company can sell the fourth car and save therefore some money.

When we discuss the results with the manager of the company, he admits that it is usually impossible to distribute the money in one day with only three cars. The difference between reality and the results of our model can be caused by several reasons:

- The operation at each office can often take more time than the estimated times in our data sets.
- The times needed for the transfer between offices were calculated for smooth operation. If a traffic jam on the way occurs, the time can be longer.
- The route planning by the model is more efficient than planning done by the employees.

To include the first two reasons in the model, we can increase both the operation times and transfer times for some percent in the data set. To assure some time reserve at each route, we can reduce the time window of each office (by moving the end of the time window to an earlier time). Another opportunity to assure some time reserve is to use all four cars. This can be assured by changing the constraint (4) from inequality to equation.

We have also tested the model on a smaller data set with only 15 offices (plus the garage) and two cars. Within an hour we have obtained an optimal solution, where only one car was used, and three offices were supplied via package.

In our future research we would like to focus on an extension of this model. We would like to add some constraints that will help with regular change of the routes according to some parameters set. Another improvement needed is to make the model solvable. We would like to try some methods based e.g. on [3].

Acknowledgements

This work was supported by the Internal Grant Agency of the University of Economics, Prague, grant No. F4/58/2017 and by the Faculty of Informatics and Statistics, University of Economics, Prague, grant no. IP400040.

References

[1] Cerny, M. (2012). *Computations, Volume III* (In Czech). Prague: Professional Publishing.
[2] Eiselt, H. A, & Sandblom, C.-L. (2000). *Integer Programming and Network Models*. Berlin: Springer-Verlag.
[3] Fabry, J., Korenar, V., & Kobzareva, M. (2011). Column Generation Method for the Vehicle Routing Problem – The Case Study. In: *Mathematical Methods in Economics 2011* (pp. 140–144). Prague: Professional Publishing.
[4] Mangiaracina, R., Song, G., & Perego, A. (2015). Distribution network design: a literature review and research agenda. *International Journal of Physical Distribution and Logistics*, 45, 506–531.
[5] Watson, M., Lewis, S., Cacioppi, P., & Jayaraman, J. (2014). *Supply Chain Network Design: Applying Optimization and Analytics to the Global Supply Chain*. New Jersey: Pearson.

Risk-sensitive and Mean Variance Optimality in Continuous-time Markov Decision Chains

Karel Sladký[1]

Abstract. In this note we consider continuous-time Markov decision processes with finite state and actions spaces where the stream of rewards generated by the Markov processes is evaluated by an exponential utility function with a given risk sensitivity coefficient (so-called risk-sensitive models). If the risk sensitivity coefficient equals zero (risk-neutral case) we arrive at a standard Markov decision process. Then we can easily obtain necessary and sufficient mean reward optimality conditions and the variability can be evaluated by the mean variance of total expected rewards. For the risk-sensitive case, i.e. if the risk-sensitivity coefficient is non-zero, for a given value of the risk-sensitivity coefficient we establish necessary and sufficient optimality conditions for maximal (or minimal) growth rate of expectation of the exponential utility function, along with mean value of the corresponding certainty equivalent. Recall that in this case along with the total reward also its higher moments are taken into account.

Keywords: Continuous-time Markov decision chains, exponential utility functions, certainty equivalent, mean-variance optimality, connections between risk-sensitive and risk-neutral models

JEL classification: C44, C61
AMS classification: 90C40, 60J10

1 Introduction

The usual optimization criteria examined in the literature on stochastic dynamic programming, such as a total discounted or mean (average) reward structures, may be quite insufficient to characterize the problem from the point of a decision maker. To this end it may be preferable if not necessary to select more sophisticated criteria that also reflect variability-risk features of the problem. Perhaps the best known approaches stem from the classical work of Markowitz (cf. [9]) on mean variance selection rules, i.e. we optimize the weighted sum of average or total reward and its variance.

In the present paper we restrict attention on undiscounted continuous-time Markov decision chains models with finite state spaces where the variability is measured either by a given non-zero value of the risk-sensitivity coefficient or in the class of long-run average optimal policies by choosing policies with minimal mean-variance. Observe that in the former case along with the total reward also its higher moments are taken into account.

The article is structured as follows. Section 2 contains notations and summary of basic facts on continuous-time Markov reward processes. The heart of the paper are sections 3 and 4. Markov models with exponential utility function (called risk-sensitive Markov chains) are studied in section 3 along with the corresponding moment generating functions. Section 4 is devoted to mean-variance optimality in continuous-time Markov decision chains; the results are mostly adapted from [17]. Section 5 summarized algorithmic procedures for finding optimal decisions. Conclusions, including comparison of reported results with similar results in the literature, are made in Section 6.

2 Notations and Preliminaries

In this note we consider Markov decision processes with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$ in continuous-time. In particular, the development of the considered Markov decision process $X = \{X(t), t \geq 0\}$ (with finite state space $\mathcal{I}$) over time is governed by the transition rates $q(j|i,a)$, for $i, j \in \mathcal{I}$, depending on the selected action $a \in \mathcal{A}_i$. For $j \neq i$ $q(j|i,a)$ is the transition rate from state i into state j, $q(i|i,a) = \sum_{j \in \mathcal{I}, j \neq i} q(j|i,a)$ is the transition rate out of state i. As concerns reward rates, $r(i)$ denotes the rate earned in state $i \in \mathcal{I}$, and $r(i,j)$ is the transition reward accrued to a transition from state i to state j.

Let $\xi(t) := \int_0^t r(X(\tau)\mathrm{d}\tau + \sum_{k=0}^{N(t)} r(X(\tau^-), X(\tau^+))$, obviously $\xi_(t)$ is the (random) reward obtained up to time t, where $X(t)$ denotes the state at time t, $X(\tau^-)$, $X(\tau^+)$ is the state just prior and after the kth jump, and $N(t)$

[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

is the number of jumps up to time t. Similarly $\xi(t',t) := \int_{t'}^{t} r(X(\tau))\mathrm{d}\tau + \sum_{k=N(t')}^{N(t)} r(X(\tau^-), X(\tau^+))$ is the total (random) reward obtained in the time interval $[t',t)$; hence $\xi_(t+\Delta) = \xi(\Delta) + \xi(\Delta, t+\Delta)$ or $\xi(t+\Delta) = \xi(t) + \xi(t, t+\Delta)$.

In this note we shall suppose that the obtained random reward, say ξ, is evaluated by an exponential utility function, say $u^\gamma(\cdot)$, i.e. utility functions with constant risk sensitivity depending on the value of the risk sensitivity coefficient γ. In case that $\gamma > 0$ (the risk seeking case) the utility assigned to the (random) reward ξ is given by $u^\gamma(\xi) := \exp(\gamma\xi)$, if $\gamma < 0$ (the risk averse case) the utility assigned to the (random) reward ξ is given by $u^\gamma(\xi) := -\exp(\gamma\xi)$, for $\gamma = 0$ it holds $u^\gamma(\xi) = \xi$ (risk neutral case). Hence we can write

$$u^\gamma(\xi) = \mathrm{sign}(\gamma)\exp(\gamma\xi) \tag{1}$$

and for the expected utility we have (E is reserved for expectation)

$$\bar{U}^{(\gamma)}(\xi) := \mathsf{E}u^\gamma(\xi) = \mathrm{sign}(\gamma)\mathsf{E}[\exp(\gamma\xi)], \quad \text{where} \quad \mathsf{E}\left[\exp(\gamma\xi)\right] = \sum_{k=0}^{\infty} \mathsf{E}\,\frac{1}{k!}(\gamma\xi)^k. \tag{2}$$

Then for the corresponding certainty equivalent $Z^\gamma(\xi)$ we have

$$u^\gamma(Z^\gamma(\xi)) = \mathrm{sign}(\gamma)\mathsf{E}[\exp(\gamma\xi)] \Longleftrightarrow Z^\gamma(\xi) = \gamma^{-1}\,\ln\{\mathsf{E}\left[\exp(\gamma\xi)\right]\}. \tag{3}$$

From (2),(3) we immediately conclude that

$$Z^\gamma(\xi) \approx \mathsf{E}\xi + \frac{\gamma}{2}\mathsf{Var}\,\xi. \tag{4}$$

In this note we focus attention on properties of the expected utility and the corresponding certainty equivalents if the stream of rewards generated by continuous-time Markov reward chain is evaluated by exponential utility functions. Recall that exponential utility function is separable, i.e. $u^\gamma(\xi^{(1)} + \xi^{(2)}) = \mathrm{sign}(\gamma)u^\gamma(\xi^{(1)}) \cdot u^\gamma(\xi^{(1)})$, what is very important for sequential decision problems. For what follows we shall need some more notions and notation.

A (Markovian) policy controlling the decision process is given as a piecewise constant right continuous function of time. In particular, $\pi = f(t)$, is a piecewise constant, right continuous vector function where $f(t) \in \mathcal{F} \equiv \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$, and $f_i(t) \in \mathcal{A}_i$ is the decision (or action) taken at time t if the process $X(t)$ is in state i. Since π is piecewise constant, for each π we can identify the time points $0 < t_1 < t_2 \ldots < t_i < \ldots$ at which the policy switches; we denote by $f^i \in \mathcal{F}$ the decision rule taken in the time interval $(t_{i-1}, t_i]$. Policy which takes at all times the same decision rule, i.e. $\pi \sim f$, is called stationary; $Q(f)$ is the transition rate matrix with elements $q(j|i, f_i)$.

Let for $f \in \mathcal{F}$ $Q(f) = [q_{ij}(f_i)]$ be an $N \times N$ matrix whose ijth element $q_{ij}(f_i) = q(j|i, f_i)$ for $i \neq j$ and for the iith element we set $q_{ii}(f_i) = -q(i|i, f_i)$. The sojourn time of the considered process X in state $i \in \mathcal{I}$ is exponentially distributed with parameter $[q(i|i, f_i)]$. Hence the expected value of the reward obtained in state $i \in \mathcal{I}$ equals $\tilde{r}_i(f_i) = [q(i|i, f_i)]^{-1}\, r(i) + \sum_{j\in\mathcal{I}, j\neq i} q(j|i, f_i)\, r(i, j)$ and $\tilde{r}(f)$ is the (column) vector of reward rates at time t.

Recall that the row sums of a transition rate matrix $Q(f)$ are equal to null. Hence $\rho(f) = 0$ is an eigenvalue of $Q(f)$, the corresponding eigenvector is a unit vector. Moreover, the real part of every other eigenvalue of $Q(f)$ is negative. In particular, if for some $i \in \mathcal{I}$ and $f \in \mathcal{F}$ it happens that $\sum_{j\neq i} q_{ij}(f_i) < q_{ii}(f_i)$ on reaching state i the process X stops with probability $q_{ii}(f_i) = -\sum_{j\neq i} q_{ij}(f_i)$ and if $Q(f)$ is irreducible then for the spectral radius of $\tilde{Q}(f) = Q(f)$ it holds $\tilde{\rho}(f) < 0$ and the real part of every other eigenvalue of $Q(f)$ is less than $\tilde{\rho}(f)$. Observe that $[I - \tilde{Q}(f)]^{-1} = \sum_{k=0}^{\infty}[\tilde{Q}(f)]^k$ exists[2]. It can be shown (see e.g. Gantmakher [1]) that the matrix $\tilde{Q}(f)$ is nonsingular.

Using policy $\pi = f(t)$ means that if the Markov chain X was found to be in state i at time t, the action chosen at this time is $f_i(t)$, i.e. the ith coordinate of $f(t) \in \mathcal{F}$. For any policy $\pi = f(t)$ the accompanying set of transition rate matrices $\{Q(f(t)), t \geq 0\}$ determines a continuous-time (in general, nonstationary) Markov process.

Let $P(\cdot,\cdot,\pi)$ be the $N \times N$ matrix of transition functions associated with Markov chain X, i.e. for each $0 \leq s \leq t$ the ijth element of $P(\cdot,\cdot,\pi)$, denoted $P_{ij}(s,t,\pi)$, is the probability that the chain is in state j at time t given it was in state i at time s and policy π is followed. Obviously, by the well-known Chapman–Kolmogorov equation $P(s,t,\pi) = \sum_{u\in\mathcal{I}} P(s,u,\pi)\, P(u,t,\pi)$ for each $0 \leq s < u < t$. The values $P(s,t,\pi)$ are absolutely continuous in t and satisfy the system of differential equations (except possibly where the piecewise constant policy switches)

$$\frac{\partial P(s,t,\pi)}{\partial t} = P(s,t,\pi)\, Q(f(t)), \quad \frac{\partial P(s,t,\pi)}{\partial s} = -Q(f(s))P(s,t,\pi) \tag{5}$$

[2]In this note the symbol I is reserved for unit matrix, similarly e is reserved for a unit column vector.

where $P(s,s,\pi) = I$. In what follows it will be often convenient to let $P(t,\pi) = P(0,t,\pi)$. By (5) we then immediately get for any $t \geq 0$

$$\frac{\mathrm{d}P(t,\pi)}{\mathrm{d}t} = P(t,\pi)\,Q(f(t)) \quad \text{along with} \quad P(t,\pi) = I + \int_0^t P(u,\pi)Q(f(u))\mathrm{d}u. \tag{6}$$

Moreover, if the considered time horizon t tends to null, i.e. if for any piecewise constant policy $\pi = f(t)$ the considered time horizon $\Delta \downarrow 0$, for the ij-th element of $P(\Delta,\pi)$ we have

$$P_{ij}(\Delta,\pi) = \begin{cases} 1 + q_{ii}(f_i(t))\Delta + o(\Delta^2) & \text{for } i = j \\ q_{ij}(f_i(t))\Delta + o(\Delta^2) & \text{for } i \neq j \end{cases} \tag{7}$$

It is well known that for any stationary policy $\pi \sim f$ there exists $\lim_{t\to\infty} P(t,\pi) = P^*(\pi)$ and, moreover, that for any $t \geq 0$ (for stationary policy $\pi \sim f$ we write f instead of π)

$$P(t,f)\,P^*(f) = P^*(f)\,P(t,f) = P^*(f)\,P^*(f) = P^*(f), \tag{8}$$

$$Q(f)\,P^*(f) = P^*(f)\,Q(f) = 0 \quad \text{where} \quad P(t,f) = \exp(Q(f)t) = \sum_{k=0}^{\infty} \frac{1}{k!}[Q(f).t]^k. \tag{9}$$

3 Exponential Utility and the Corresponding Moments

Supposing that stationary policy $\pi \sim f$ is followed, let $U_i^{(\gamma)}(t,f) := \mathsf{E}_i^f \mathrm{e}^{\gamma\xi(t)}$ and by (2) $\bar{U}_i^{(\gamma)}(t,f) := \mathsf{E}_i^f u^\gamma(\xi(t))$ be the expected value of the random reward evaluated according the exponential utility function $u^\gamma(\cdot)$ earned up to time t if the process starts in state i (obviously $\bar{U}_i^{(\gamma)}(t,f) = \mathrm{sign}(\gamma)\,U_i^{(\gamma)}(t,f)$). Recall that $\mathsf{E}_i^f[\exp(\gamma\xi(t))] = U_i^{(\gamma)}(t,f)$ is also the moment generating function of $\xi(t)$. Hence (cf. (2)) for the k-th moment of $\xi(t)$ it holds

$$M_i^{(k)}(\xi(t),f) := \frac{\mathrm{d}^k}{\mathrm{d}\gamma^k}\mathsf{E}_i^f[\exp(\gamma\xi(t)))]|_{\gamma=0} = \mathsf{E}_i^f\,\xi(t))^{k-1} \tag{10}$$

and the Taylor expansion around $\gamma = 0$ reads

$$U_i^{(\gamma)}(\xi(t),f) = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} M_i^{(k)}(\xi(t),f) \;\text{ for }\; |\gamma| < h. \tag{11}$$

Since $u^\gamma(\cdot)$ is separable and multiplicative we have

$$u^\gamma(\xi(t)+\Delta) = \mathrm{sign}(\gamma)\,u^\gamma(\xi(\Delta))\cdot u^\gamma(\xi(\Delta,t)), \quad \text{or} \quad u^\gamma(\xi(t+\Delta)) = \mathrm{sign}(\gamma)\,u^\gamma(\xi(t))\cdot u^\gamma(\xi(t,t+\Delta)).$$

On taking expectations we immediately conclude that for $U_i^{(\gamma)}(t,f) := \mathsf{E}_i^f\{\mathrm{e}^{\gamma\xi(t)}\}$ we have (δ_{ij} is the Kronecker symbol)

$$U_i^{(\gamma)}(t+\Delta,f) = \sum_{j=1}^{N} P_{ij}(\Delta,f)\cdot[\mathrm{e}^{\gamma r(i)\Delta}\delta_{ij} + \mathrm{e}^{\gamma r(ij)}(1-\delta_{ij})]\cdot U_j^{(\gamma)}(t,f). \tag{12}$$

Since

$$\mathrm{e}^{\gamma r(i)\Delta} = 1 + \gamma r(i)\Delta + o(\Delta^2), \quad P_{ij}(\Delta,f) = \begin{cases} 1 + q_{ii}(f_i)\Delta + o(\Delta^2) & \text{for } i = j \\ q_{ij}(f_i)\Delta + o(\Delta^2) & \text{for } i \neq j \end{cases}$$

on letting $\Delta \to 0+$ we conclude that

$$U_i^{(\gamma)}(t+\Delta,f) = (1+q_{ii}(f_i)\Delta)\mathrm{e}^{\gamma r(i)\Delta}U_i^{(\gamma)}(t,f) + \sum_{j=1,\,j\neq i}^{N} q_{ij}\Delta\mathrm{e}^{\gamma r(ij)}\cdot U_j^{(\gamma)}(t,f) + o(\Delta^2). \tag{13}$$

Since the process X is time homogeneous and $(1+\gamma r(i)\Delta)(1+q_{ii}(f_i)\Delta) = 1 + (q_{ii}(f_i) + \gamma r(i))\Delta + o(\Delta^2)$ after some manipulation we arrive at

$$\frac{\mathrm{d}U_i^{(\gamma)}(t,f)}{\mathrm{d}t} = (q_{ii}(f_i) + \gamma r(i))U_i^{(\gamma)}(t,f) + \sum_{j=1,j\neq i}^{N} q_{ij}(f_i)\mathrm{e}^{\gamma r(ij)}\cdot U_j^{(\gamma)}(t,f) \tag{14}$$

that can be also written in matrix form as

$$\frac{\mathrm{d}U^{(\gamma)}(t,f)}{\mathrm{d}t} = \bar{Q}^{(\gamma)}(f)\cdot U^{(\gamma)}(t,f) \tag{15}$$

where $U^{(\gamma)}(t,f) = [U_i^{(\gamma)}(t,f)]$ is a (column) vector, and $\bar{Q}^{(\gamma)}(f) = [\bar{q}_{ij}^{(\gamma)}(f_i)]$ is an $N\times N$ matrix with nonnegative off-diagonal elements

$$\bar{q}_{ij}^{(\gamma)}(f_i) = \begin{cases} q_{ii}(f_i)+\gamma\cdot r(i) & \text{for } i=j \\ q_{ij}(f_i)\cdot \mathrm{e}^{\gamma\, r(i,j)} & \text{for } i\neq j \end{cases}$$

Hence by (15) if $U^{(\gamma)}(0,f) = e$

$$U^{(\gamma)}(t,f) = \exp[\bar{Q}^{(\gamma)}(f).t]\cdot e = \sum_{k=0}^{\infty}\frac{1}{k!}[\bar{Q}^{(\gamma)}(f).t]^k\cdot e. \tag{16}$$

To study asymptotic behavior of $U^{(\gamma)}(t,f)$ first recall (cf. Gantmakher [1]) that for any matrix with non-negative off-diagonal elements there exists a positive eigenvalue (called the dominant eigenvalue) such that the real part of any other eigenvalue is nongreater than the dominant eigenvalue. In addition, the corresponding eigenvector can be selected positive and even strictly positive if the considered matrix is irreducible or at least unichain. In particular, let $\sigma^{(\gamma)}(f)$ be the dominant eigenvalue of $\bar{Q}^{(\gamma)}(f)$ and $v^{(\gamma)}(f)$ the corresponding eigenvector, i.e.

$$\bar{Q}^{(\gamma)}(f)\cdot v^{(\gamma)}(f) = \sigma^{(\gamma)}(f)\cdot v^{(\gamma)}(f).$$

Observe that if the matrix $Q(f)$ is unichain then $\bar{Q}^{(\gamma)}(f)$ remains unichain at least if risk-sensitive coefficient γ is sufficiently close to null.

4 Mean Variance Optimality

Another approach how to handle variability-risk features in continuous-time Markov decision processes is based on mean variance selection rules, i.e. we optimize the weighted sum of average or total reward and its variance. To this end, we focus attention on undiscounted continuous-time Markov decision chains, i.e. we assume that the risk sensitive coefficient $\gamma = 0$, hence the corresponding utility $u^{(0)}(\xi) = \xi$ and the expected utility $\bar{U}^{(0)}(\xi) := \mathsf{E}\xi$. Hence for the selected long-run stationary policy the variability is measured either by the ratio of long-run average reward and long run average variance. Of course, it is optimal in the class of policies with a fixed average reward to select policy with minimal average variance. This method that transforms finding minimal variance to finding minimal average reward of suitably transformed Markov decision process was initially suggested in [7] for discrete-time Markov decision chains, similar procedure for continuous-time models were used in [3],[4],[17].

To begin with, we start with well-known approach for evaluating and finding optimal average policy for continuous-time Markov decision processes (see e.g. [5, 2]). Similar approach can be also used for finding second moment and the corresponding variance of the considered Markov reward chain.

Considering the risk-neutral models, on taking expectations we conclude that for $U_i^{(0)}(t,f) := \mathsf{E}_i^f\{\xi(t)\}$ we have

$$U_i^{(0)}(t+\Delta,f) = r(i)\cdot\Delta + (1+q_{ii}(f_i)\cdot\Delta)U_i^{(0)}(t,f) + \Delta\sum_{j=1,j\neq i}^{N} q_{ij}(f_i)[r(ij)+U_j^{(0)}(t,f)] + o(\Delta^2) \tag{17}$$

Since the process X is time homogeneous, after some manipulations, on letting $\Delta\to 0+$, we arrive at

$$\frac{\mathrm{d}U_i^{(0)}(t,f)}{\mathrm{d}t} = r(i) + \sum_{j=1,j\neq i}^{N} q_{ij}(f_i)\cdot r(ij) + \sum_{j=1,j\neq i}^{N} q_{ij}(f_i)\cdot[U_j^{(0)}(t,f) - U_i^{(0)}(t,f)] \tag{18}$$

It is well-known from the literature if the considered continuous-time Markov chain has a single class of recurrent states that the long-run average reward is independent on the initial state and $g(f) = \lim_{t\to\infty}\frac{1}{t}U_i^{(0)}(t,f)$ exists and is independent of the starting state. In addition, it is well-known that there exist vector $w(f)$ (with elements $w_i(f)$) such that

$$U_i^{(0)}(t,f) = g(f)\cdot t + w_i(f) - [P(t,f)w(f)]_i \qquad ([w]_i \text{ denotes the } i\text{th entry of the column vector } w).$$

In what follows we shall be also interested in the second moment and the corresponding variance of random reward $\xi(t)$. Since $\mathsf{E}_i^f[\xi(t+\Delta)]^2 = \mathsf{E}_i^f[\xi(t)]^2 + 2\cdot\mathsf{E}_i^f\xi(t)\cdot\Delta + \mathsf{E}_i^f[\Delta]^2$, for the second moment of $\xi(t)$, say $S_i(f)(\xi) := \mathsf{E}_i^f(\xi)^2$, we get

$$\begin{aligned} S_i(t+\Delta,f) &= (1+\Delta\cdot q_{ii}(f_i))\left\{2\Delta\cdot r(i)U_i^{(0,f)}(t,f)+S_i(t,f)\right\} \\ &\quad +\Delta\sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)\left\{[r(ij)]^2+2r(ij)U_j^{(0)}(t,f)+S_j(t,f)\right\}+o(\Delta^2). \end{aligned} \tag{19}$$

Hence (similarly to (18) for $\Delta\to 0+$ from (18) we get

$$\begin{aligned} \frac{\mathrm{d}S_i(t,f)}{\mathrm{d}t} &= 2r(i)U_i^{(0)}(t,f)+\sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)\left\{[r(ij)]^2+2r(ij)U_j^{(0)}(t,f)\right\} \\ &\quad +\sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)[S_j(t,f)-S_i(t,f)]. \end{aligned} \tag{20}$$

Since $\sigma_i(t,f) = S_i(t,f) - [U_i^{(0)}(t,f)]^2$ we thus obtain:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\sigma_i(t,f) &= \frac{\mathrm{d}}{\mathrm{d}t}S_i(t,f) - 2U_i^{(0)}(t,f)\frac{\mathrm{d}}{\mathrm{d}t}U_i^{(0)}(t,f) \\ &= 2r(i)U_i^{(0)}(t,f)+\sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)\left\{[r(ij)^2+2r(ij)U_j^{(0)}(t,f)\right\}+\sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)[S_j(t,f)-S_i(t,f)] \\ &\quad -2U_i^{(0)}(t,f)\left\{r(i)+\sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)r(ij)+\sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)[U_j^{(0)}(t,f)-U_i^{(0)}(t,f)]\right\} \end{aligned} \tag{21}$$

By substituting $S_j(t,f) = \sigma_j(t,f) + [U_j^{(0)}(t,f)]^2$ (21) can be rewritten as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_i(t,f) = \sum_{j\in\mathcal{I}} q_{ij}(f_i)\sigma_j(t,f) - 2U_i^{(0)}(t,f)\left\{\sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)r(ij)+\sum_{j\in\mathcal{I}} q_{ij}(f_i)U_j^{(0)}(t,f)\right\} \tag{22}$$

Using a more detailed analysis (see [17]) it can be shown that

$$G(f) := \lim_{t\to\infty}\frac{1}{t}\sigma_i(t,f) \tag{23}$$

exists and is independent of the initial state i. For unichain models the values $G(f)$ of the average variance can be calculated as average reward of a continuous time Markov chain where one stage reward in state i, say $s_i(f)$, is equal to

$$s_i(f) = \sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)\{[r(ij)+w_j(f)]^2-[w_i(f)]^2\}+2[r(i)-g(f)]w_i(f), \text{ or} \tag{24}$$

$$s_i^{(1)}(f) = \sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)\{[r(ij)+w_j(f)]^2-[w_i(f)]^2\}+2r(i)w_i(f), \text{ or} \tag{25}$$

$$s_i^{(2)}(f) = \sum_{j\in\mathcal{I},j\neq i} q_{ij}(f_i)\{[r(ij)]^2+2r(ij)\}+2r(i)w_i(f). \tag{26}$$

5 Finding optimal control policy

Considering continuous-time Markov decision processes, it has been shown in section 3 that the growth of risk-sensitive optimality is governed by matrices with nonnegative off-diagonal elements. Recalling that for any matrix with nonnegative off-diagonal elements there exists a real eigenvalue (called the dominant eigenvalue) such that the real part of any eigenvalue of the matrix is nongreater than the dominant eigenvalue, for maximizing the risk-sensitive average reward it suffices to find stationary policy defining feasible matrix with maximum possible real eigenvector. To this end we can employ policy iteration or value iteration specified in [14]. This methods also enable to generated upper and lower estimates on the maximum possible dominant eigenvalue in each iteration step.

Another approach for finding optimal policies was studied in section 4. Using this approach the decision maker selects by standard policy or value iteration methods the set of all (non-randomized) stationary policies with maximal average reward. In the next step in the class of stationary policies maximizing the average reward the decision maker selects the policy with minimal average variance. Observe that the construction of the policy with minimal average variance is limited only on the class of optimal policies (resp. another selected class of policies) since the formulas for construction the policy with minimal variance heavily employ values of some coefficients constructed for optimal policies (cf. formulas (24)–(26)). Up to now we looked for policies with minimal variance in the class of policies with maximal total reward. However, considering stationary policy (even randomized), not maximizing the average reward, policy and value iteration methods can be used for finding policies guaranteeing the desired reward with minimal possible variance.

6 Conclusions

In this note we studied risk-sensitive average optimality in risk-sensitive continuous-time Markov decision chains. The obtained results extend some older result reported for uncontrolled models in [13] and are overlapping with recent results reported in [18] obtained by different methods and are similar to the results reported in [3] and [4].

The obtained result are also confronted with so-called mean-variance optimality criterion for finding policies with minimal average variance in certain class of policies (usually in the class of policies with maximal average reward). The results reported in section 4 are similar to Thm. 10.5 in [2], or Thm.3.4 in [10] (in these references no transient rewards are considered).

Acknowledgement

This research was supported by the Czech Science Foundation under Grant 18-02739S.

References

[1] Gantmakher, F.R. (1959). *The Theory of Matrices.* London: Chelsea.

[2] Guo, X. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes: Theory and Applications.* Berlin: Springer.

[3] Guo, X. and Song, X. (2009). Mean-variance criteria for finite continous-time Markov decision processes. *IEEE Trans. Automat. Control* 54, 2151–2157.

[4] Guo, X., Ye, L. and Yin, G.A. (2012). Mean-variance optimization problem for discounted Markov decision processes. *European J. Oper. Res.* 220, 423–429.

[5] Howard, R.A. (1960). *Dynamic Programing and Markov Processes.* New York: Wiley.

[6] Howard, R.A. and Matheson, J. (1972). Risk-sensitive Markov decision processes. *Manag. Science* 16, 356–369.

[7] Mandl, P. (1971). On the variance in controlled Markov chains. *Kybernetika* 7, 1–12.

[8] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance* 7, 77–92.

[9] Markowitz, H. (1959). *Portfolio Selection - Efficient Diversification of Investments.* New York: Wiley.

[10] Prieto-Rumeau, T. and Hernández-Lerma, O. (2009). Variance minimization and the overtaking optimality approach to contin uous-time controlled Markov chains. *Math. Method Oper. Res.* 70, 527–540.

[11] Puterman, M.L. (1994). *Markov Decision Processes – Discrete Stochastic Dynamic Programming.* New York: Wiley.

[12] Ross, S.M. (1983). *Introduction to Stochastic Dynamic Programming.* New York: Academic Press.

[13] Sladký, K. (2008). Risk-sensitive discrete- and continuous-time Markov reward processes. In: M. Reiff (Ed.), *Proceedings of the International Scientific Conference Quantitative Methods in Economics (Multiple Criteria Decision Making XIV)*(pp. 272-281). Bratislava: University of Economics.

[14] Sladký, K. (2008). Growth rates and average optimality in risk-sensitive Markov decision chains. *Kybernetika* 44, 205–226.

[15] Sladký, K. (2013). Risk-sensitive and mean variance optimality in Markov decision processes. *Acta Oeconomica Pragensia* 7, 146–161.

[16] Sladký, K. (2017). Second order optimality in Markov decision chains. *Kybernetika* 53, 1086–1099.

[17] Van Dijk, N.M. and Sladký, K. (2006). The variance of discounted Markov decision processes. *J. Appl. Probab.* 43, 1044–1052.

[18] Wei, Q. and Chen, X. (2016). Continuous-time Markov decision processes under risk-sensitive average cost criterion. *Oper. Res. Lett.* 44, 457–462.

Efficiency Comparison in the Development of Building Projects Sector

Michaela Staňková[1], David Hampel[2]

Abstract. This article focuses on the estimation of technical efficiency of the selected companies within the development of buildings projects sector in the Czech Republic. Data envelopment analysis method was selected to calculate the efficiency. We use annual accounting data of the companies that are available in the Amadeus database of Bureau van Dijk. Material costs, costs of employees, capital and assets are selected as input variables. Operating revenue (turnover) and added value represents output variables. The dataset contains financial data of 46 companies in three periods. Efficiency rating models are compiled for constant returns to scale as well as for variable returns to scale. This makes it possible to evaluate the area of the scale efficiency. All models are radial input oriented and constructed separately for each of the three monitored periods. Recommendations are given to selected companies to remedy their situation.

Keywords: company accounting data, data envelopment analysis, efficiency, linear programming.

JEL Classification: C44, C67
AMS Classification: 90B50, 90C08

1 Introduction

Prior to 1950, economists believed that companies are always making effective use of their resources in order to make a profit. Later did the concept of efficiency be associated with the use of resources in such a way as to maximize production. Farrell's work has been a major shift in efficiency field, see [5]. Since then, researchers have begun to distinguish between technical and allocative efficiency.

Two different approaches are used to measure efficiency: the parametric and the non-parametric approach. Parametric methods rely on statistical techniques to estimate the parameters of a production function. This approach is criticized mainly for these two points: the assumption that the production function has the same functional form for all the companies; the fact that econometric estimation of efficiency can produce inconsistent parameter estimates. Non-parametric methods compare the observed inputs and outputs of each firm with that of the most performing firms in the information dataset without having information about the production function. The data envelopment analysis (DEA) method is the most used one from the non-parametric methods for evaluating relative efficiency. The basis of the DEA method is based on Farrell's work [4]. Later, different models were developed on this basis. The three most used DEA models are: the CCR model (with constant returns to scale assumption) by Charnes, Cooper, and Rhodes [10]; the BCC model (that considers variable returns to scale) by Banker, Charnes, and Cooper [2] and the Additive model (which combines both inputs and outputs orientations) by Charnes et al. [9].

This article focuses on the comparison of relative technical efficiency in the selected construction sector. The construction industry is an important (and traditional) part of the Czech economy as well as a significant component of the GDP. This sector is very different from other sectors of the national economy. The specifics of the building sector are the especially seasonal nature of production, depending on the natural conditions, the mobility of place of production etc. This means that companies in this sector has to deal with constantly changing external factors.

Due to the nature of our dataset, the DEA method is suitable for the purpose of evaluating efficiency. This method is widely used to analyze efficiency in sectors with homogeneous units, for example the banking (like [7] and [8]) and healthcare (see [11] and [12]). Another suitable area for this method is construction industry. Recent studies in this industry are [6] and [14]. Development of building projects sector is a sector with strong

[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

competition that is caused by the struggle for customers as well as limited natural resources. Companies have similar technology available and have the same social and legal environment within the Czech Republic.

2 Material and Methods

The financial (annual accounting) data are collected from the Amadeus database of Bureau van Dijk. We deal with companies within the development of buildings projects sector (NACE code 411) from the Czech Republic with data available in years 2013, 2014 and 2015. A total of 46 companies were included in the efficiency evaluation. With regard to the theory, we choose six characteristics representing both production and cost characteristics often used for efficiency assessment, see Table 1. The DEA models are constructed separately for particular years. In Table 1, there are basic characteristics of financial variables used in our analysis.

	Char.	Capital (Input)	Costs of employees (Input)	Material costs (Input)	Total assets (Input)	Added value (Output)	Operating revenue (Output)
2013	Min	3.64	3.54	0.84	221.57	-782.44	1.46
	Median	72.9	147.65	53.34	4971.59	293.52	1405.97
	Max	364559.97	4192.22	3682.01	203800.65	30986.25	124121.08
2014	Min	3361	0.79	0.47	303.79	-2799.76	13.20
	Median	72314	108.47	54.99	7366.51	355.38	1272.80
	Max	360786341	4423.58	4656.64	201382.92	61721.13	143098.71
2015	Min	0.00	0.74	1.15	234.55	-142.60	0.00
	Median	74.00	107.16	51.54	5833.09	401.89	1312.11
	Max	370089.74	4939.23	9370.13	260636.31	40011.1	160917.14

Table 1 Minimum, maximum and median of each selected variable in thousands EUR in individual years

The input CCR models with the constant returns to scale (CRS) assumption and also input BCC model assuming variable returns to scale (VRS) were used for the efficiency evaluation. Based on input (X) and output (Y) variables it can be measured the input oriented efficiency of each company *o* by solving *n* times the following CCR model:

$$\min_{\theta,\lambda} \theta \qquad (1)$$
$$\theta x_o \geq X\lambda$$
$$Y\lambda \geq y_o$$
$$\lambda \geq o.$$

The input-oriented BCC model also evaluates the efficiency of company *o* ($o = 1,\ldots,n$), the only difference with the CRS model (1) is the addition of the condition

$$\sum_{j=1}^{n} \lambda_j = 1. \qquad (2)$$

This procedure makes it possible to evaluate the area of the scale efficiency. Scale efficiency can be expressed as the ratio of the CRS efficiency to the VRS efficiency. Efficiency under VRS is pure technical efficiency. Efficiency under CRS is overall technical efficiency – it is the combined effect of scale and pure technical efficiency. Scale efficiency measures the proportion VRS efficient output levels relative to what output would company have if this unit moves to most productive scale size.

To calculate the technical efficiency, a radial measurement method was used. To obtain the possible input excesses and output shortfalls, the second stage problem, that incorporates calculation of the optimal value of CRS/VRS and corrects radial inefficiency, was solved. In these models the additional input excesses and output shortfalls can be defined by so-called slack vectors to obtain feasible solution. We use the Malmquist Productivity Index presented for example in [3] to represent the results of company's progress or decline in different years, similarly to [13]. Technical details about all above stated models can be found in [1]. All calculations were performed in computational system Matlab R2017b.

3 Results

Technical efficiency results under CRS and VRS during the whole time period are summarized in Table 2. Efficiency scores based on models under VRS show that more than half of the companies in this selected industry are not working effectively. In the case of CRS efficiency, the number of effective companies is decreasing. The largest decline is visible in 2015, when the number of effective companies within the CRS is reduced by more than half in comparison with VRS. Although it is not possible to judge trends for just three years, it is worthwhile to notice that the number of least-performing and fully effective companies is relatively stable.

It was found that only four companies are labeled as effective at all three periods and all DEA models. Namely, these companies are Moravská stavební – INVEST, a.s., Portland Trust, s.r.o., Benti-byty, s.r.o. a K–STAVBY, a.s. A common factor for all these companies is, in particular, long-term functioning on the market. Portland Trust, s.r.o. has been operating in the Czech Republic since 1997. It can be assumed that management experience has been reflected in the resulting efficiency values. Besides, these companies operate mainly in Prague and Brno. These large cities give them more opportunities in a relatively small area.

	2013		2014		2015	
Efficiency	**CRS (%)**	**VRS (%)**	**CRS (%)**	**VRS (%)**	**CRS (%)**	**VRS (%)**
[0; 20)	23.91	2.17	21.74	4.35	19.57	4.35
[20; 40)	21.74	10.87	21.74	15.22	19.57	17.39
[40; 60)	6.52	15.22	13.04	17.39	13.04	6.52
[60; 80)	13.04	8.70	13.04	6.52	13.04	8.70
[80; 1)	8.70	15.22	2.17	8.70	13.04	17.39
1	26.09	47.83	28.26	47.83	21.74	45.65

Table 2 Distribution of companies' efficiency under CRS and VRS in all three periods

Figure 1 shows the situation of all companies[3] with regard to their resulting efficiency score under CRS and VRS. The circle symbol represents results of 2013, the square symbol represents results of 2014 and the triangle symbol denotes the results of 2015. Companies whose final score is projected near the green line are characterized by a high level of scale efficiency. If the resulting score of the enterprise is moving away from this line, the scale efficiency is decreasing.

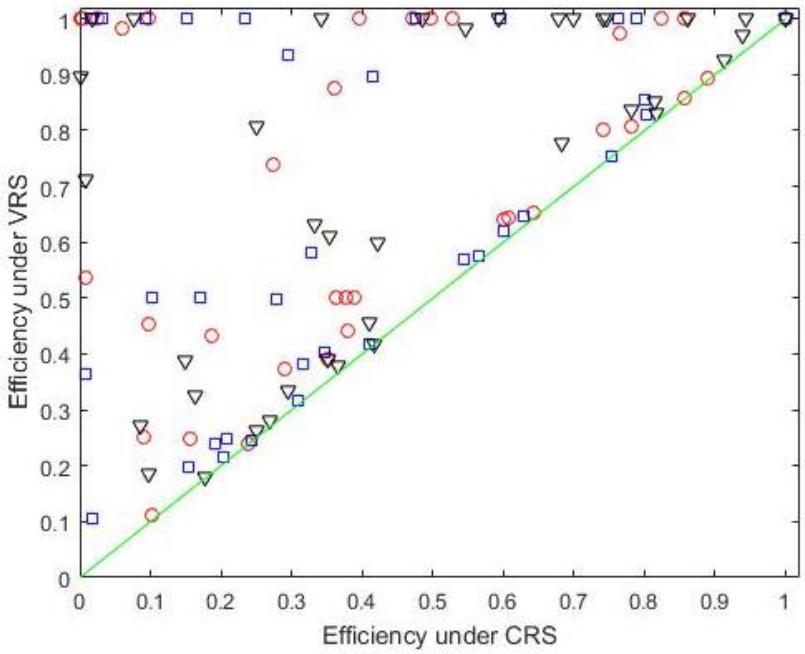


Figure 1 Efficiency score under CRS and VRS in all periods

[3] Companies with the same results – for example, the fully efficient company, which represents point [1; 1] – overlap each other.

The Scale inefficiency is due to the company dimension. The biggest scale inefficiency can be found with companies that find themselves on the graph at the top left. Its size of operations is not optimal so that modifications on its size can render the unit more efficient. The division of companies in the picture is relatively stable. In 2014, companies that undertake high or full efficiency in the DEA model under VRS, but have low scale efficiency, are much more separated.

Figure 2 represents the development of 46 company's efficiency in 2014. Based on Malmquist Productivity Index it can be generally said that companies achieving values higher than one achieved technical efficiency improvements. Values of less than one indicate deterioration in efficiency (i.e. greater inefficiency).

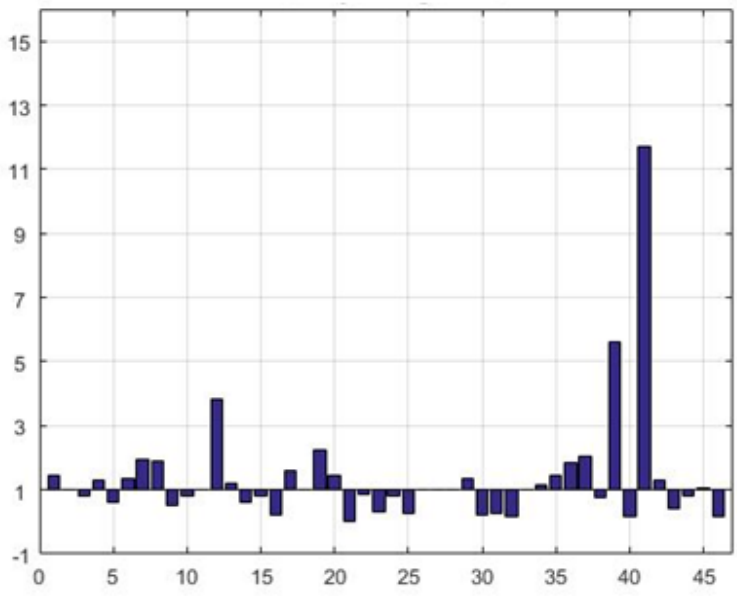


Figure 2 Companies efficiency score under CRS and VRS in 2014

Figure 3 shows the efficiency change in 2015. Based on the general results coming from Figure 2 and 3, it is obvious that some companies have improved over the period under review, but others are getting worse. In each period, at least one company rapidly improved its effectiveness - for example, OXES, s.r.o. (number 41) in 2014, CZ STAVEBNÍ HOLDING, a.s. (number 21) in 2015. This rapid change is caused by several factors. For example, in the case of CZ STAVEBNÍ HOLDING, a.s., the following factors are involved: recovery after the crisis, completion and sale of several major projects and, last but not least, a merger of a purpose-built company, which led a separate accounting.

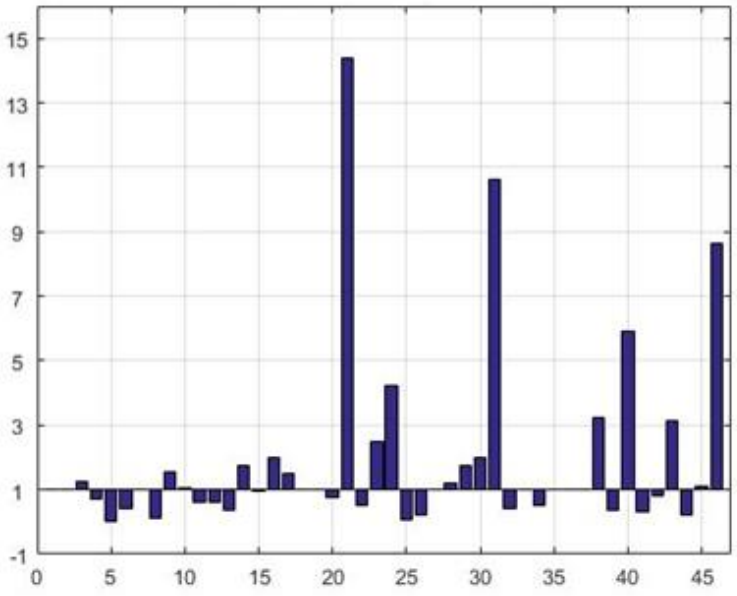


Figure 3 Companies efficiency score under CRS and VRS in 2015

As mentioned above, most companies have been identified as inefficient by the model. It was also found that most of these inefficient businesses, even after proportional reductions in inputs, cannot reach the efficiency frontier (i.e. to its efficient target). In these cases, slacks are needed to push the inefficient companies to the frontier. Table 3 represents the percentage of inefficient companies and inefficient companies with slacks in all three periods.

	Model	2013	2014	2015
Percentage of inefficient companies related to all companies (%)	CRS	73.91	71.74	78.26
	VRS	52.17	52.17	54.35
Percentage of inefficient companies with slacks related to all inefficient companies (%)	CRS	70.59	84.85	80.56
	VRS	87.50	91.67	88.00

Table 3 Percentage of inefficient companies and inefficient companies with slacks in all three periods

Some companies just need to significantly reduce (or increase) one input (output) to remove the slack. In general, change is needed in two or more areas. In Table 4, there are frequencies of slacks in individual variables. The most common problem was the company's capital. The capital structure is usually planned with regard to the company's strategy. According to the Ministry of Industry and Trade, the construction sector is following a relatively unbalanced financial structure with a slight upward trend on the side of liabilities as financial sources in 2014. Companies with a more conservative funding strategy are therefore considered ineffective in the DEA models.

	Model	Capital (Input)	Costs of employees (Input)	Material costs (Input)	Total assets (Input)	Added value (Output)	Operating revenue (Output)
2013	CRS	18	6	4	5	11	2
	VRS	10	5	8	4	7	6
2014	CRS	17	4	13	5	13	9
	VRS	9	2	9	3	6	11
2015	CRS	18	12	1	4	12	6
	VRS	11	9	3	4	9	9

Table 4 Number of slacks in individual models in all three periods

4 Discussion

This article has produced interesting results in the field of company efficiency in the Development of building projects sector in Czech Republic. Unfortunately, there are no studies to compare the results for the sector so specified in the Czech Republic or other countries. For a more insight into the inefficiency of these companies, it would be appropriate to extend the period under review. Some companies in this sector have long-term contracts that exceed one or more accounting periods. This makes analysis of efficiency more complicated – because it is not easy to see the amount of unsold unfinished projects in the financial statements. This information can often be included in annual company reports. However, this information (if available) needs to be searched outside the Amadeus database and individually for each company. In addition, with an appropriate period setting, DEA Window analysis could bring interesting results.

5 Conclusion

In this paper, we compared technical efficiency of companies in the development of buildings projects sector in three periods. It was found that most units in this industry were identified as inefficient companies by the DEA model (under CRS as well as VRS). Their size of operations is not optimal; modifications on its size can render the inefficient company more efficient. However, most of these inefficient units, even after the proportional reductions in inputs, cannot reach the efficiency frontier. In these cases, slacks are needed to push these inefficiency companies to the frontier. Most often, companies were lagging behind in terms of capital and added value. Besides, it was found that the ratio of effective and inefficient companies is relatively stable over the reporting period.

Acknowledgements

This article was supported by the grant No. PEF/DP/2018027 of the Grant Agency IGA PEF MENDELU.

References

[1] Álvarez, I. C., Barbero, J., Zofío, J. L. (2016). *A Data Envelopment Analysis Toolbox for MATLAB, Working Papers in Economic Theory 2016/03*, Universidad Autónoma de Madrid (Spain): Department of Economic Analysis. Available at: (https://ideas.repec.org/p/uam/wpaper/201603.html).

[2] Banker, R., Charnes, A., Cooper, W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science,* 30 (9), 1078–1092.

[3] Cooper, W. W., Seiford, L. M., Tone, K. (2007). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software*. 2nd edition. New York: Springer Science & Business Media.

[4] Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society,* 120 (3), 253–290.

[5] Førsund, F. (2018). Economic interpretations of DEA. *Socio-Economic Planning Sciences,* 61, 9–15.

[6] Hu, X., Liu, Ch. (2015). Managing undesirable outputs in the Australian construction industry using Data Envelopment Analysis models. *Journal of Cleaner Production,* 101, 148–157.

[7] Huang, T., Lin, C., Chen, K. (2017). Evaluating Efficiencies of Chinese Commercial Banks in the Context of Stochastic Multistage Technologies. *Pacific-Basin Finance Journal,* 39, 68–77.

[8] Haque, F., Brown, K. (2017). Bank ownership, regulation and efficiency: Perspectives from the Middle East and North Africa (MENA) Region. *International Review of Economics& Finance,* 39, 68–77.

[9] Charnes, A., Cooper, W. W., Golany, B., Seiford, L., Stutz, J. (1985) Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. Journal of Econometrics 30 (1–2), 91–107.

[10] Charnes, A., Cooper, W., Rhodes, E. (1987). Measuring the efficiency of decision making units. *Management science,* 2 (6), 429–444.

[11] Khushalani, J. Ozcan, Y. A. (2017). Are hospitals producing quality care efficiently? An analysis using Dynamic Network Data Envelopment Analysis (DEA). *Socio-Economic Planning Sciences,* 60, 15–23.

[12] Mitropoulos, P., Talias, M. A., Mitropoulos, I. (2015). Combining stochastic DEA with Bayesian analysis to obtain statistical properties of the efficiency scores: An application to Greek public hospitals. *European Journal of Operational Research,* 243 (1), 302–311.

[13] Rashidi, S. F., Barati, R., Kamali, Z. (2014). Measurement Efficiency by Malmquist Method in Data Envelopment Analysis. *Trends in Applied Sciences Research,* 9, 54–60.

[14] Xue, X., Wu, H., Zhang, X., Dai, J., Su, Ch. (2017). Measuring energy consumption efficiency of the construction industry: the case of China. *Journal of Cleaner Production,* 107, 509–515.

A simulation analysis of the accuracy of median estimators for different sampling designs

Tomasz Stachurski[1]

Abstract. In the paper the problem of median estimation is considered. The median is a parameter of practical interest, especially in socio-economic research. There are analyzed different estimation strategies. We consider three median estimators including estimator that does not use any auxiliary information, a ratio median estimator and also a proposition of a synthetic ratio estimator, that is applicable for any sampling design and can be used in small area estimation even though, the domain sample size equals zero. Furthermore, three different sampling designs are studied including a proposition of the sampling design, which is based on a distance measure between values of auxiliary variable and its median. In the design-based simulation study, the properties of proposed estimator and sampling design are considered.

Keywords: median, bootstrap, survey sampling, Monte Carlo.

JEL Classification: C83
AMS Classification: 62D05

1 Introduction

Statistics is the field of science that has applications in many areas. Statistical methods are widely used in socio-economic research. One of the problems that is considered in such studies is the measurement of the central tendency. There are a lot of measures of central tendency - the most popular and widely discussed in literature is mean. However, mean is not always preferable as a typical parameter, because it can be strongly influenced by outliers – very small masses of points that situated long distances from the bulk of the distribution (see [1] p. 178). Some alternative of the measure of location is the median.

Let x_0 to be the point that divided the population into to equal parts, so that half of the population elements have values smaller or equal x_0, whereas the other half have values greater or equal x_0. Then x_0 is called a median of the distribution (see [1] p. 178). It is worth pointing out that every distribution has at least one median, but it is not always uniquely determined. Let consider the population distribution function $F(y)$, which can be defined as a proportion of population elements not exceeding y_k what can be written as (see [6] p. 197):

$$F(y) = \frac{1}{N}\left(\#A_y\right) \tag{1}$$

where $A_y = \{i: i \in \Omega \ \wedge y_i \leq y\}$ and by $\#A_y$ we denote the number of elements in the set A_y. If there is only one intersection of the cumulative distribution function $F(y)$ with straight line $y = \frac{1}{2}$ then we have determined case and the median is uniquely defined. It is also possible that $F(y)$ has and line $y = \frac{1}{2}$ have a whole closed interval in common. In that indeterminate case, each point in this common interval is a median.

In order to estimate a median using only sample data $y_i \in s$ in [6] pp. 197-198, there is shown a general procedure consisting of two steps:

- estimation the cumulative distribution function $\hat{F}(y)$;
- estimation the median as $\widehat{M} = \hat{F}^{-1}(0{,}5)$.

What is more, that procedure can be generalized to estimate any quantile of a distribution (median is the 50% quantile).

Nowadays, researches are not only interested in population characteristics, but in some distinguished subpopulation parameters (with small or even zero sample sizes) as well. Inference about subpopulation is especially useful in allocation some financial resources e.g. to ensure equitable growth for regions (see [2] p. 5]. That is why, in the paper there are presented methods of the median estimation in the population and subpopulations. Those methods belongs to the discipline called ‘small area estimation’, where small area is defined as “any domain for which using direct estimators (that uses information about studied variable only from the domain of interest) cannot produce estimates with satisfactory level of accuracy (see [5] p. 2).

[1] University of Economics in Katowice, Department of Statistics, Econometrics and Mathematics, 50, 1 Maja Street, Katowice, Poland , email: tomasz.stachurski@ue.katowice.pl

The main purposes of this paper are:

- presenting a proposition of the synthetic ratio estimator of the domain median;
- introducing a sampling design using a distance measure between values of an auxiliary variable and its median;
- a simulation-based analysis of the properties of the proposed estimator and sampling designs in comparison to other median estimators under simple and complex sampling design.

Obtained results can be valuable for any practitioners who are interested in evaluating the average level of studied variable, especially when some economic variables are studied, which usually are asymmetrically distributed, what implies that using mean as the parameter of location is not appropriate.

2 Estimation a population and subpopulation median

Let define a sampling design. It is a function $p(s)\colon S \to [0,1]$ based on sample space S, which assigns a probability of drawing to each sample. Moreover, this function has to fulfill two conditions:

- probability of drawing each sample must be non-negative:

$$\bigwedge_{s\in S} p(s) \geq 0, \tag{2}$$

- sum of probabilities for all samples s form sample space S is equal to 1:

$$\sum_{s\in S} p(s) = 1. \tag{3}$$

The r-order inclusion probability is a probability that units denoted by $i_1, i_2, \dots i_r$ belong to the sample, what can be written as follows:

$$\pi_{i_1,i_2,\dots i_r} = \sum_{\{s:i_1,i_2,\dots,i_r\in s)} p(s). \tag{4}$$

In [6] pp. 199-200, there is proposed a general procedure of estimation a median that is applicable of any sampling design. The first step is to arrange in non-decreasing order all of the sample elements y_i as follows:

$$y_1 \leq y_2 \leq \cdots \leq y_n \tag{5}$$

and the corresponding inclusion probabilities π_i:

$$\pi_1, y_2, \dots, \pi_n. \tag{6}$$

Inclusion probabilities π_i can be obtained using any sampling design (also complex sampling designs), which are studied in more detailed in section 3.

Let define cumulative sums that are necessary to estimate the population median. Set $B_0 = 0$ and then:

$$B_1 = \frac{1}{\pi_1}; \tag{7}$$

$$B_2 = \frac{1}{\pi_1} + \frac{1}{\pi_2}; \tag{8}$$

and so on that:

$$B_l = \sum_{i=1}^{l} \frac{1}{\pi_i}. \tag{9}$$

Hence, $B_n = \sum_i^n \frac{1}{\pi_i} = \widehat{N}$ and the median estimator presented in [6] p. 200 can be written as (10):

$$\widehat{M} = \begin{cases} y_{l,} \text{ if } B_{l-1} < 0{,}5\widehat{N} < B_l \\ \frac{1}{2}(y_l + y_{l+1}) \text{ if } B_l = 0{,}5\widehat{N} \end{cases}. \tag{10}$$

Estimator (10) does not use any additional information about population units. It is very common, that there is available information about auxiliary variable for the whole population. Using such auxiliary information can lead to increase of the estimation accuracy (provided that auxiliary variable is highly correlated with variable under study).

A natural generalization of a widely known ratio estimator of the mean or the total is the ratio estimator for the median, which was firstly proposed in [4]. It was applicable for simple sampling designs. Simple modification of it presented in [8] p. 86 enables to use that ratio estimator for any sampling design. The ratio estimator of domain median is given by formula (11):

$$\widehat{M}_{\Omega_d}^{R} = \frac{\widehat{M}_{y_{\Omega_d}}}{\widehat{M}_{x_{\Omega_d}}} M_{x_{\Omega_d}}, \tag{11}$$

where $\widehat{M}_{y_{\Omega_d}}$ and $\widehat{M}_{x_{\Omega_d}}$ are obtained using estimator (10) and $M_{x_{\Omega_d}}$ is the median of auxiliary variable X in dth domain (it is assumed that values of X are known for all the population elements). What is more, if in (11) we assume that $\Omega_d = \Omega$, then we obtain the estimator of the population median.

Ratio estimator given by (11) is a direct estimator. It means, that it is based on the information about variable under study only from the domain of interest (see [5] p. 1). That implies that (11) cannot be used for a domain where sample size is zero. Hence, it is proposed ratio synthetic estimator of a domain median which is given by formula (12):

$$\widehat{M}_{\Omega_d}^{SR} = \frac{M_{x_{\Omega_d}}}{\widehat{M}_{x_\Omega}} \widehat{M}_{y_\Omega}, \tag{12}$$

where $\widehat{M}_{y_\Omega}, \widehat{M}_{x_\Omega}$ are the estimators of the population median using (10) and $M_{x_{\Omega_d}}$ is the known value of the population median of the auxiliary variable. Estimator (12) is a domain indirect estimator, because it uses information about the study variable also from domains another than domain of interest. Synthetic ratio estimator given by (12) can be also obtained using ratio estimator (for the population) given by formula (11) as follows:

$$\widehat{M}_{\Omega_d}^{SR} = \frac{M_{x_{\Omega_d}}}{\widehat{M}_{x_\Omega}} \widehat{M}_{y_\Omega} = \frac{M_{x_{\Omega_d}}}{\widehat{M}_x} \widehat{M}_y \cdot \frac{M_{x_\Omega}}{M_{x_\Omega}} = \frac{M_{x_{\Omega_d}}}{M_{x_\Omega}} \frac{\widehat{M}_y}{\widehat{M}_x} M_{x_\Omega} = \frac{M_{x_{\Omega_d}}}{M_{x_\Omega}} \widehat{M}_\Omega^R. \tag{13}$$

Even though synthetic ratio estimator is biased estimator of the median, it can be profitable to use it instead of direct domain estimators, because usually variances of direct estimators in domains (especially when domain sample size is small) are much more than variances of indirect estimators. Moreover, indirect estimators, e.g. synthetic ratio estimator of domain median given by (12) can be used even when domain sample size is zero.

3 Algorithms of sampling

There are a lot of different sampling designs. The basic division is based on equality or inequality first order inclusion probabilities in those sampling designs. In the paper there are studied two basic kinds of sampling designs:

- simple random sampling without replacement, where first-order inclusion probabilities are given by (14):

$$\pi_i = \frac{n}{N} \tag{14}$$

where n is the sample size and by N it is denoted the population size;

- sampling proportional to size where first-order inclusion probabilities are given by (15)

$$\pi_i = \frac{n x_i}{\sum_{i \in \Omega} x_i}. \tag{15}$$

Sampling proportional to size it is profitable when the aim of the study is estimation the total value (in order to avoid underestimation the parameter). In our case, it is preferable to select units, which are as close as possible to the median of auxiliary variable. If the variable under study is highly correlated with the auxiliary variable, then we suppose that observations which values of the auxiliary variable are close to its median, their values of the studied variable should be also close it median of that variable. That is why, we propose to use sampling proportional to a distance measure between values of auxiliary variable and its median as follows:

$$z_i = \frac{1}{|x_i - M_x| + c}, \tag{16}$$

where c is a positive constant determined by a researcher so as to avoid division by zero.

Hence, inclusion probabilities are obtained using

$$\pi_i = \frac{nz_i}{\sum_{i\in\Omega} z_i}. \tag{17}$$

It is worth noting, that variable z takes higher value for units that are close to the median and lower values that are located long distances from the median. As a result it is more probable to sample elements that are located close to the median of an auxiliary variable.

4 Estimation the variance of median estimators

In order to measure the precision of an estimator, its variance must be calculated. In our study some bootstrap methods were used in order to estimate the variance of the presented in section 2 median estimators. Algorithm of bootstrap was proposed in [3]. This method has a lot of application including constructing confidence intervals, hypothesis testing and estimation the variance of an estimators (see [7] pp.145-147).

In the paper there are considered two algorithms of estimation the variance using the bootstrap technique:

1. Bootstap for the independent variables. This algorithm consists of the following steps:
   a) draw B times sample with replacement from the original sample s;
   b) for each of B bootstrap sample, calculate the parameter of interest $\hat{\theta}_b^*$ in the same way as estimator $\hat{\theta}$ was calculated for the original sample;
   c) the observed distribution of $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ can be treated as an estimation of sampling distribution of estimator $\hat{\theta}$ and its variance can be obtained using formula (18)

$$V(\hat{\theta}) = \frac{1}{B-1}\sum_{i=1}^{B}\left(\hat{\theta}_b^* - \hat{\theta}^*\right)^2, \tag{18}$$

where $\hat{\theta}^* = \frac{1}{B}\sum_i^B \hat{\theta}_b^*$.

2. Bootstrap for complex sampling designs proposed in [6] pp. 443-444. It is a modification of procedure for independent variables, that takes into account inclusion probabilities of each element, that is why that algorithm is applicable for any sampling design. The procedure consists of the following steps:
   a) construct artificial population denoted by Ω^* composed of replicates of the sampled units. For each of the unit from the sample, we create $\frac{c}{\pi_i}$ elements in the artificial population Ω^*. Due to the fact that quotients $\frac{c}{\pi_i}$ does not always be integer, the constant c must be determined so that those quotient were approximately integer (see [9] p. 88);
   b) draw resamples from the population Ω^* (with replacement) with probabilities $p_i = \frac{\pi^*}{n}$ for $i \in \Omega^*$;
   c) then follows steps (b)-(c) from the procedure 1.

5 Simulation study

In the section there are presented the results of the design-based simulation study, which was conducted using R. The aim of the study was analysis of the properties of considered median estimators (11)-(13) under three sampling designs: sampling proportional to size (denoted on Figures (1)-(5) by PPS), sampling proportional to a distance measure (denoted by PPD) given by (16) and simple random sampling without replacement (denoted by SWOR). For considered sampling strategies in the simulation study were analyzed:

- relative biases of median estimators;
- relative root of the mean square (RMSE) error of estimators;
- relative biases of variances of the median estimators.

The simulation study was conducted both for real and generated data. We use real data on municipalities in Sweden (dataset available in [6] pp. 652-69). The study variable was revenues from municipal tax in 1985 (in millions of kronor) and the auxiliary variable was the number of municipal employees in 1984. The correlation coefficient between those variables is above 0,99. From this dataset we remove three outliers, because they could have extremely strong influence on the results of the simulation study, especially if the number of iterations is not big enough. After removing those extreme values, both studied variables are strongly positively skewed. Moreover, we consider artificial data generated from the multidimensional normal distribution. We assume the same level of the correlation between variables and the same mean as in real dataset.

In the first step, we analyzed relative biases of the median estimators (see Figure 1 and Figure 2). For the generated data for each of considered sampling strategies, values of simulation biases of domain and population median estimators were very close to zero. For the real data, more differences between estimators occurred. The

median estimator given by (10) is highly biased, in one case the relative bias exceeds 100%. It is also worth noting, that using the proposed in the paper sampling design, where inclusion probabilities are given by (17) led to much decrease of relative biases comparing to other sampling designs. Ratio median estimator given by (11) and the proposed synthetic ratio estimator (12) has much better properties. Their relative biases ranges from about -10% up to 10%. It means that using ratio estimators of the median lead to significant decrease in relative biases in case of strongly skewed variables.

Secondly, we analyzed accuracy of the median estimators by calculating their relative root mean square error (RMSE). For the generated data, for each sampling strategy RMSE were small – in most cases values of relative RMSE were below 0,4%. But the best results were obtained for the proposed synthetic ratio estimator. For the real dataset much greater values of relative RMSE were received. For the median estimator given by (10) relative RMSE in some domains were above 100%. Ratio estimators (given by (11) and (12)) are great alternative, because using them lead to substantial decrease of relative RMSE. For those ratio estimators, rRMSE does not exceed 25%. Due to the fact, the synthetic ratio estimator is intended for estimation domain median, then on figures (1)-(4) there are not presented its biases or RMSE in the population.

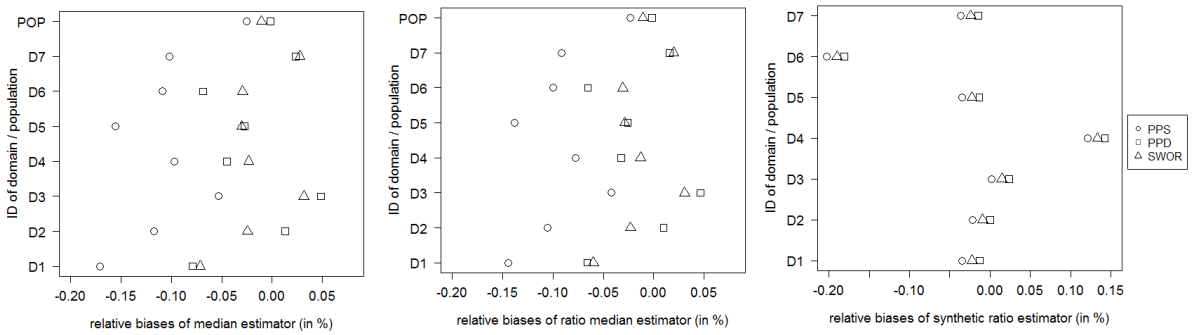


Figure 1 Relative biases of the median estimators for the artificial data

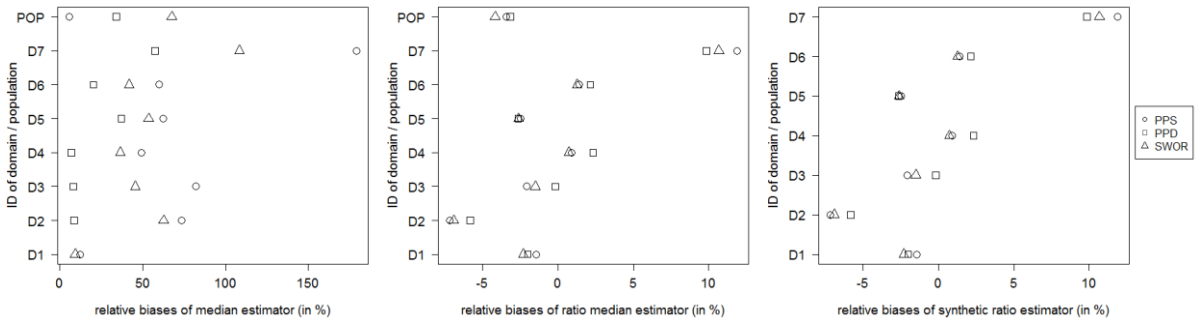


Figure 2 Relative biases of the median estimators for the real data

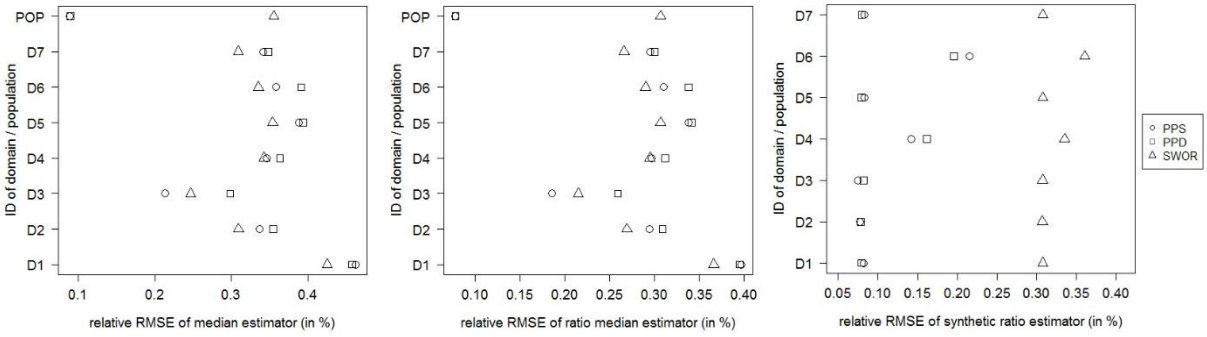


Figure 3 Relative RMSE of the median estimators for the artificial data

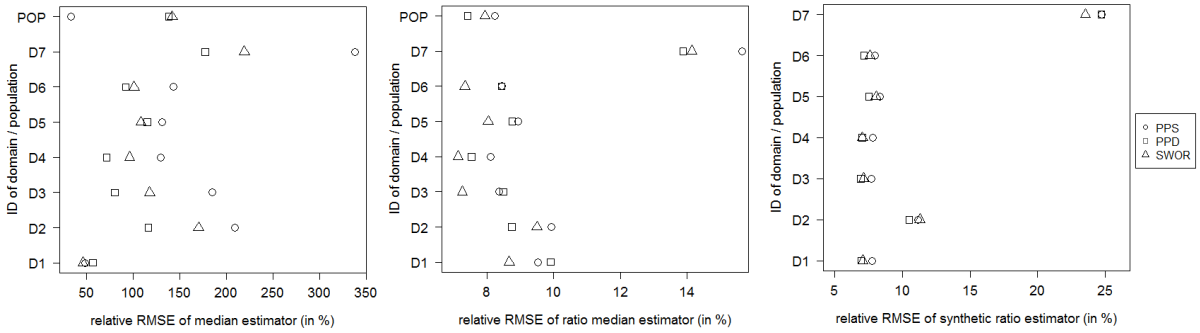


Figure 4 Relative RMSE of the median estimators for the real data

We also compared in table 1. relative biases of variances of median estimators estimated by bootstrap for independent variables and estimated by bootstrap for complex sampling designs. It means that estimation the variance using bootstrap for complex sampling designs has better properties.

ID of subpopulation / the population	Bootstrap – technique I	Bootstrap – technique II
1	-37,27%	-65,56%
2	64,03%	222,7%
3	83,77%	-40,85%
4	-50,09%	53,94%
5	-47,99%	-71,11%
6	-62,48%	-31,05%
7	113,6%	-33,45%
the population	-94,45%	-83,16%

Table 1 Relative biases of estimators of design-variance of the median ratio estimator for the real data

6 Conclusion

In the paper we conduct a comprehensive simulation-based analysis of properties of median estimators for simple and complex sampling strategies. We can conclude for the asymmetric variables (what is very common, especially in economic research) using ratio median estimators including the proposed synthetic ratio estimators leads to sustainable decrease of relative biases and relative root mean square error comparing to the estimator that does not use auxiliary information. Moreover, the proposed sampling design based on a distance between values of the auxiliary variable and its median can be profitable especially when we analyze the estimator which estimator does not use auxiliary information. The simulation study also shown that using bootstrap technique for complex sampling design in order to estimate the variance of median estimators is more accurate that bootstrap for independent variables.

References

[1] Cramér H.: *Mathematical Methods of Statistics*. Asia Publishing House, Bombay, 1962.
[2] Domański Cz., and Pruska K.: *Metody statystyki małych obszarów*. Wydawnictwo UŁ, Łódź, 2001.
[3] Efron B.: Bootstrap Methods: Another Look at the Jacknife. *Annals of Statistics*, no 7, 1979, 1-26.
[4] Kuk Y.C.A., and Mak T.K.: Median estimation in the Presence of Auxiliary Information. Jounral of the Royal Statistical Society. Series B (Methodological), Vol. 51, No. 2 (1989), 261-269.
[5] Rao J.N.K., and Molina I.: *Small Area Estimation*. John Wiley & Sons, New York, 2015.
[6] Särndal C.E, Swensson B., and Wretman J.: *Model Assisted Survey Sampling*. Springer-Verlag, New York, 1992.
[7] Stachurski T.: O estymacji wartości średniej za pomocą przedziałów ufności Neymana oraz metod bootstrapowych. In: *Zastosowania metod ilościowych w badaniach ekonomicznych* (Żądło T., eds.). Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, 2014, 143-150.
[8] Thompson M.E., and Godambe V.P.: Estimating Functions and Survey Sampling. In: *Handbook of Statistics. Sample Surveys: Inference and Analysis, Vol. 29B* (Pfeffermann D., and Rao C.R., eds.), Elsevier Science, Oxford, 2009.
[9] Wywiał J.: Wprowadzenie do metody reprezentacyjnej. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, 2010.

The Issue of Weights for Constructing the Composite Index of Social Inequalities

Michaela Staníčková [1], Jiří Franek[2]

Abstract. Measurement of progress plays a crucial role in improving the prosperity in any territories. Several methods of evaluating the level of territorial development exist, most of the methods have their own limitations. There is no uniform methodological approach to the evaluation of territorial development and composites indices (CIs) are one of the suitable techniques that allow to aggregate and summarize the multidimensional characteristics. CIs are very common for benchmarking the mutual and relative progress of territories in a variety of policy domains. The proliferation of the production of CIs by all the major international organizations is a clear symptom of their importance in policy-making. Constructing CIs includes several steps that must be made and corresponding methods have to be chosen. Construction of CIs includes not only a problem of selection of indicators and quantitative method (normalizing methods and aggregation formulas) of their processing but especially a question of weights. The paper deals with the selection of relevant tool for determining the weights having a significant impact on the overall CI for measuring the social inequalities and benchmarking framework of the EU NUTS 2 regions. The paper also explores the drawbacks of CIs, and questions whether these CIs can adequately serve as policy-setting mechanisms.

Keywords: Composite index, European Union, NUTS 2 region, social inequalities, weights

JEL Classification: C43, C82, R11, R12
AMS Classification: 28D20, 62H25, 90B50

1 Introduction

If we look at the EU evolution over the past decades, substantial progress has been made in terms of building an internal market and an economic and monetary union, albeit not without problems, as the 2008 crisis has shown. It looks, as if the EU and its Member States were mostly thinking in economic terms, hoping that economic solutions will fix all social problems at once. To negate the importance of social issues is to undermine the EU foundations (Allmendinger, Driesch, 2014). 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. The low growth performance in the EU over the recent decades has increased concerns regarding an increasing economic dispersion, income inequality at large, and social exclusion. The recent economic crisis revealed many of the weaknesses of the current European economic policy, not least at the level of its fiscal policy, monetary policy, industrial policy, and social policy, and its inability to address problems related to inequality. Several factors can explain this widespread interest; especially the revival of growth theory (Romer, 1990; Aghion, Howitt, 1998) was contemporaneous with literature on economic convergence (Sala-i-Martin, 2006; Quah, 1997).

The level of social inequalities belongs to important indicators influencing the socio-economic development and other processes taking place in the social and economic realm (Melecký, 2015). It should be mentioned that such multidimensional phenomena as income disparity and poverty might be analysed from many different perspectives, including the national and international, also within the EU. In this case the measurement of progress plays a crucial role in improving the prosperity. There exist several methods of evaluating the level of territorial development. However, they have their drawbacks. There is no uniform methodological approach to the evaluation of territorial development. The composites indices (CIs) are one of the suitable techniques that allow to aggregate and summarize the multidimensional characteristics. CIs are very common for benchmarking the mutual and relative progress of territories in a variety of policy domains. Constructing CIs includes several steps that must be made, and corresponding methods have to be chosen. Construction of CIs includes not only a problem of selection

[1] 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.
[2] VŠB-TU Ostrava, Faculty of Economics, Department of Business Administration, Sokolská třída 33, 702 00 Ostrava 1, Czech Republic, e-mail: jiri.franek@vsb.cz.

of indicators and quantitative method (normalizing methods and aggregation formulas) of their processing but especially a question of weights. This paper focuses on the issue of selection of the relevant tool for determining the weights that have a significant impact on the overall CI for measuring the social inequalities and benchmarking framework of the EU NUTS 2 regions – the case of Visegrad Four (V4) countries. The paper also explores the drawbacks of CIs, and questions whether these CIs can adequately serve as policy-setting mechanisms.

2 Methodological background of weighting in composite indices

In general terms, an indicator is a quantitative or a qualitative measure derived from a series of observed facts that can reveal relative positions in a given area. Indicators are pieces of information that summarize the characteristics of a system or highlight what is happening in a system. CIs are useful in their ability to integrate large amounts of information into easily understood formats and are valued as a communication and political tool. They are often a compromise between scientific accuracy and the information available at a reasonable cost. However, the construction of CIs suffers from many methodological difficulties (Nasierowski, 2016). There are several steps to be followed in constructing CIs, such as (1) developing a theoretical framework; (2) identifying and developing relevant indicators; (3) assessing the quality of dataset and assessing the relationships among indicators; (4) standardizing indicators to allow comparisons; (5) weighting indicators and aggregation indicators; (6) testing for robustness and sensitivity. The quality of the aggregation convention is, yet, an important ingredient to guarantee the consistency between the assumptions used and the ranking obtained. Indeed, the overall quality of CI depends crucially on the way how mathematical model is embedded in the evaluated process. This is especially true for the choice of weights that remains the more important source of uncertainty and debate (Freudenberg, 2003). Composite indicators are much like mathematical or computational models. As such, their construction falls under universally accepted scientific rules for encoding. The European Commission adopts a definition type of composite indicator developed in this chapter, i.e. a composite indicator based on sub-indicators that have no common meaningful unit of measurement and there is no obvious way of weighting them (Saisana, Tarantola, 2002, p. 5). Weighting and aggregation systems have a crucial effect on the outcome of each composite index. There is not only one proper method. Generally, a composite index CI can be understood as weighted linear aggregation of a set of variables, see e.g. Munda, Nardo (2005) or Freudenberg (2003):

$$CI = \sum_{i=1}^{N} w_i X_i \tag{1}$$

where X_i is normalized variable, w_i is weight of variable X_i where $\sum_{i=1}^{N} w_i = 1$ and $0 \le w_i \le 1,\ i = 1, 2, ..., N.$

Although various functional forms for the underlying aggregation rules of a composite indicator have been developed in literature, in standard practice a weighted linear aggregation rule applied to a set of variables. The evaluation of the criteria weights may be subjective, objective and integrated. The list of the most common weighting methods is summarised, for instance, in OECD (2008).

2.1 Subjective weighting

Essentially the decision maker can select to use equal weights. This imply that all indicators or sub-indices in the composite index have equal importance on the performance being measured. In many composite indices, equal weights are used for reasons of simplicity. With the equal weighting approach, there is the risk that certain performance aspects will be double weighted (Freudenberg, 2003). Moreover, the results of an equally weighted index are still conditioned by the choice of the normalisation method. The most prominent substantive justification for equal weighting goes back to Occam's razor. Since it is probably impossible to obtain agreement on weights, the simplest arrangement is the best choice. Stated differently, equal weighting may sometimes not even be an adequate description of the debate in composite indices construction (Saisana, Tarantola, Schulze, Laurens, Moesen, Puyenbroeck, 2005, p. 30). Explicit weighting includes mostly three main methods: experts' opinion, survey weighting (public opinion) and Analytic Hierarchic Process (AHP), see e.g. Sharpe, Andrews (2012), OECD (2008), Freudenberg (2003), Saisana, Tarantola (2002). Experts' opinion is based on subjective personal judgement. In survey weighting, people are asked to express their concern about certain problems measured by the indicators. The simplest way to replicate valuations of society relatively accurately is to collect a representative sample of the population. However, in many cases, there are no resources to conduct such a survey. Therefore, survey weighting remains the optimal solution to the weighting issue, but it is rarely achievable (Sharpe, Andrews, 2012, p. 5). AHP is one of the most popular analytical techniques for complex decision-making problems. AHP enables to handle decision situations involving subjective judgments, multiple decision makers and to provide measures of consistency of preference. The AHP creates a hierarchical structure of the criteria. The weights of indicators within AHP can be determined by the pairwise comparison (Saaty's method).

2.2 Statistical/objective weighting

Although explicit weights have the favourable quality of being transparent, the options available often do not satisfy all necessary conditions. Given that the weights are chosen based on patterns in the data, the researchers cannot be criticized for a bias in their assignment of weights (Sharpe, Andrews, 2012, p. 5). The objective methods determine criteria weights by solving mathematical models automatically without any consideration of the decision. As common methods are considered e.g. Entropy method, CRITIC (The Criteria Importance through Intercriteria Correlation), literature review, mean weight, standard deviation, statistical variance procedure, see e.g. Minarčíková (2016). These are based on information on alternatives can be used only in case of a finite number of alternatives (Ginevičius, Podvezko, 2005). The data uncertainty can be captured using the entropy method. It requires knowledge of the values of all the criteria for all variants in the matrix R. In the theory of information, the entropy is the criterion of uncertainty posed by a discrete probability distribution pi. This degree of uncertainty is expressed by [3] in the formula:

$$S(p_1, p_2, ..., p_n) = -c.\sum_{i=1}^{n} p_i.ln\, p_i, \tag{1}$$

where c is a positive constant. Equation (1) express entropy in statistical concept, therefore entropy can be found as probability distribution p_i and thus terms of entropy and probability are considered as synonyms. Suppose all p_i equal, then for given i, $p_i = \frac{1}{n}$ reaches $S(p_1, p_2, ..., p_n)$ maximum value. From matrix **R** we can determine share of the i-th variant on the sum of the j-th criteria for all criteria p_{ij} from the formula:

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^{p} r_{ij}}, i = 1,2,...,p, j = 1,2,...,k. \tag{2}$$

For the j-th criterion entropy (s_j) can be determined by formula:

$$s_j = -c.\sum_{i=1}^{p} p_{ij}.ln\, p_{ij}, j = 1,2,...,k. \tag{3}$$

If suppose $c = \frac{1}{\ln p}$, then $0 \le s_j \le 1$ is guaranteed. Non normalized entropy weight of j-th criteria (d_j) can be found in formula:

$$d_j = 1 - s_j, j = 1,2,...,k, \tag{4}$$

while respective normalized weights w_i are obtained from the formula:

$$w_j = \frac{d_j}{\sum_{j=1}^{k} d_j}, j = 1,2,...,k. \tag{5}$$

Given the availability of data for composite indices measures and their form, the entropy methods is appropriate for weights estimation further into the paper.

3 Empirical study

An empirical study for determining the weights having a significant impact on the overall CI for measuring the social inequalities and benchmarking framework of the EU NUTS 2 regions was performed. The selection of adequate indicators of social disparities observed at the level of NUTS 2 regions has been identified within the Reports on Economic, Social and Territorial Cohesion, i.e. the EU Cohesion reports that evaluate the trends of disparities and cohesion in the EU Member States and their NUTS 2 regions (see European Commission, 2007, 2010, 2014, 2017). The paper contribution to regional disparities in social inequality measurement is represented by the construction of the Index of Social Disparities (ISD) and the Weighted Aggregated Index of Social Disparities (WAISD) respectively, as presented in Figure 1. The procedure of ISD and WAISD is based on a combination of selected multivariate mathematical and statistical methods that lead to a unique model that includes an index of social disparities that can summarise a view of regional disparities in social inequality and more easily interpret than a set of many separate indicators, as presented in Table 1. ISD and WAISD respectively reduce the visible size of a selected set of regional social indicators without dropping the underlying information base.

Input data analysis: Collection of convenient indicators of regional disparities for 35 NUTS 2 regions of the V4 countries » Data normalisation (Z-score transformation) » Dataset of normalised variables for selected V4 35 NUTS 2 regions
Calculation of synthetic indices of disparities: Calculation of distance (Euclidean Distance Method) » Calculation of synthetic indices of disparities » Calculation of relative weight for each dimension of disparities » Calculation of a Weighted Synthetic Index of Social Disparities » Descriptive characteristics of synthetic index variability
Results and discussion: Comparison of regional disparities across the social dimension of disparities » Derivation of the cohesion level in explored NUTS 2 regions » Interpretation of results and discussion

Table 1 Basic Scheme of the Empirical Analysis
Source: authors` proposal and elaboration, 2018

Based on Table 1, in the first layer of the model, the method of a standardised variable (Z-score transformation) and method of distance from an imaginary point, presented as the square Euclidean distance from the median, is used. In the second layer, the exploratory factor analysis for a partial calculation of factor loadings (saturation) is used. Factor loadings present the correlation coefficients between the original variable and the extracted factor from the Principal Component Analysis and show how much of the variability of the factor is explained. Factor loadings therefore represent a full explanation of the role of each character (variable) in the definition of the factor. Factor loadings obtained from the factor analysis play key role in the second layer of the construction of the composite index. They are used as normalised weights of standardised individual indicators of social disparities. The normalised factor loadings for each indicator within the social dimension of disparities are therefore included in the calculation of the Index of Social Disparities (ISD). This procedure is recommended e.g. by Munda, Nardo (2005) or OECD (2008). The Synthetic Index of Social Disparities is, from a statistical point of view, designed as a modified weighted squared Euclidean distance, defined by formula (6):

$$ISD_{r,t} = \sum_{d=1}^{j} \sum_{i=1}^{k} zw_{i_S} D^{*}_{ES}(zx_{i,r,t}, z\tilde{x}_{i,r}), \tag{6}$$

where:

$ISD_{r,t}$ Index of social disparities for dimension *d-th*and region *r-th*in time *t*;

zw_{i_S} normalised weight based on factor loadings for dimension *d-th* (social) and indicator *i-th*;

$D^{*}_{ES}(zx_{i,r,t}, z\tilde{x}_{i,r})$ modified square Euclidean distance of indicator *i-th*for region *r-th*in time *t*;

$zx_{i,r,t}$ standardised value of indicator *i-th* for region *r-th*in time *t*;

$z\tilde{x}_{i,r}$ median of indicator *i-th*for region *r-th*in the whole time period;

R region; r = {1 = CZ01, … , 8 = CZ08, 9 = HU10, ..., 15 = HU33, … , 16 = PL11, …, 31 = PL63, 32 = SK01, …, 35 = SK04};

d dimension of disparities; d = {social};

i indicator of disparities; i = {1= ER15to64, …, 8 = AAP};

t time; t = {2000;…,2016}.

Weighting and aggregation systems have a crucial effect on the outcome of each composite index. There is not only one proper method. That is why this part of constructing the composite index is the most discussed and criticised by the opponents of composite indices. Although various functional forms for the underlying aggregation rules of a composite indicator have been developed in literature, in standard practice, the composite indicator as well as WAISD can be considered as a weighted linear aggregation rule applied to a set of variables as shown in Table 1. The evaluation of the criteria weights may be subjective, objective and integrated. The list of the most common weighting methods is summarised, for instance, in OECD (2008). In the concept of WAISD we used the objective approach based on the entropy method to determine the entropy weight (ew_i) for the Index of Regional Disparities. Calculation of WAISD is based on a weighted linear aggregation defined by formula (7):

$$WAISD_{r,t} = ISD_{r,t} ew_{i_S} \tag{7}$$

where:

$WAISD_{r,t}$ Weighted aggregated index of social disparities for region *r-th* in time *t*;

$ISD_{r,t}$ Index of social disparities for region *r-th* in time *t*;

ew_{i_S} normalised entropy weight for the social dimension of disparities;

r region; r = {1 = CZ01, … , 8 = CZ08, 9 = HU10, ..., 15 = HU33, … , 16 = PL11, …, 31 = PL63, 32 = SK01, …, 35 = SK04};
t time; t = {2000 …, 2016}.

The case study of regional disparities measurement and evaluation is based on 8 selected indicators of social inequalities. The selection of adequate indicators of social disparities observed at the level of NUTS 2 regions has been identified within the Reports on Economic, Social and Territorial Cohesion, i.e. the EU Cohesion reports that evaluate the trends of disparities and cohesion in the EU Member States and their NUTS 2 regions (see European Commission, 2007, 2010, 2014, 2017). The reference period 2000-2016 is determined by the selection of all indicators and their data availability for 35 NUTS 2 regions of V4 countries. The used selected indicators are social disparities as follows in Table 2.

Indicator of social disparities	Abbreviation	Criterion	Source
Employment rate	ER15to64	Maximum	Eurostat
Employment rate of women	ERw15to64	Maximum	Eurostat
Employment rate of older workers	ER55to64	Maximum	Eurostat
Unemployment rate	UR15to64	Minimum	Eurostat
Unemployment rate of youth	URy15to24	Minimum	Eurostat
Long-term unemployment	LtUR	Minimum	Eurostat
Population aged 25-34 with tertiary education attainment	PATE	Maximum	Eurostat
Annual average population change	AAP	Maximum	Eurostat

Table 2 Selected Indicators of Regional Disparities for V4 NUTS 2 Regions
Source: Eurostat, 2018; authors` proposal and elaboration, 2018

Most individual results obtained from the computed synthetic index based on the dataset of 8 selected disparities indicators revealed that the development in 35 NUTS 2 Visegrad Four regions indicates a positive trend of social disparities recorded in the reference period 2000-2016. The results in the social dimension of disparities indicate that most computed standardized values of the synthetic index converge to the optimal value (i.e. to 0) more at the end of the reference period (2016) than at the beginning of the reference period (2000). The following Table 2 shows the results of computed median values of the index of social disparities in the form of a Weighted Aggregate Index of Social Disparities (WAISD) within 35 V4 NUTS 2 regions for the whole reference period. WAISD values show that the rate of regional disparities in NUTS 2 regions with the agglomeration of capital cities (CZ01, HU10, PL12 and SK01) is rather smaller than in the rest of V4 NUTS 2 regions of each country. Based on the results of the Synthetic Index of Social Disparities, the smallest value of ISD is represented by Czech NUTS 2 regions CZ01 (Praha, 1st position), CZ02 (Střední Čechy, 3rd position) and CZ03 (Jihozápad, 4th position). There is also the Slovak region SK01 (Bratislavský kraj, 2nd position) and Hungarian HU10 (Közép-Magyarország, 7th position), and Polish PL12 (Mazowieckie, 12th position) with the smallest value of social disparities. Based on the analysis of the results, the initial presumption, that in NUTS 2 regions of the V4 countries which will be evaluated as the areas with the lowest level of disparities and highest derived level of development potential, the agglomeration of capital cities will be located, has been confirmed. Table 2 illustrates the WAISD scores division among the individual 35 NUTS 2 regions of V4 as median values, and sorts regions based on their scores into ranks.

Rank of regions by the median of WAISD								
1.	CZ01	0.024	13.	PL21	2.782	25.	PL43	4.224
2.	SK01	0.138	14.	SK02	2.924	26.	PL22	4.485
3.	CZ02	0.524	15.	PL31	2.938	27.	HU33	4.768
4.	CZ03	0.628	16.	PL11	2.966	28.	PL61	4.831
5.	CZ06	1.101	17.	PL34	2.981	29.	PL62	4.985
6.	CZ05	1.122	18.	HU21	3.057	30.	PL42	5.516
7.	PL12	1.426	19.	PL63	3.126	31.	HU23	5.714
8.	CZ07	1.609	20.	PL41	3.207	32.	SK03	5.830
9.	HU10	2.041	21.	PL33	3.744	33.	HU32	7.524
10.	HU22	2.308	22.	PL32	3.771	34.	SK04	7.727
11.	CZ04	2.493	23.	PL51	3.895	35.	HU31	7.953
12.	CZ08	2.545	24.	PL52	4.006			

Table 2 WAISD Results for V4 NUTS 2 Regions – Rank and Scores
Source: authors` calculation and elaboration. 2018

Conclusion

The measurement and evaluation of regional disparities in the social dimension within this chapter has been performed through the construction of a weighted synthetic index as an example of composite indicators calculated from standardized values of social disparities by the modified square Euclidean distance and Exploratory Factor analysis. The main advantage of the used approach lies namely in the ability to summarize the different units of measure under one synthetic characteristic (index). which is a dimension less figure. The analysis showed that, for the most part, there was a consensus in the trends of V4 countries NUTS 2 regions in terms of the attainment level of disparities and development potential. depending on the level of existing social disparities. The construction of the synthetic index and calculation of social disparities showed that since the year 2000 positive social development has been monitored in NUTS 2 regions of the Visegrad Four countries and thus the level of social cohesion recorded an increasing trend thanks to a mostly decreasing volume of regional disparities. In spite of the narrowing rate of social disparities and convergence process in the level of cohesion. significant regional disparities between the V4 countries remain.

Acknowledgements

Supported by the grant No. 17-23411Y of the Czech Science Agency and the Operational Programme Education for Competitiveness – Project No. CZ.1.07/2.3.00/20.0296.

References

[1] Aghion. P. & Howitt. P. (1998). *Endogenous growth theory*. Cambridge: MIT Press.
[2] Allmendinger. J. & Driesch. E. (2014). Social Inequalities in Europe: Facing the challenge. *Discussion Paper P 2014-005*. Berlin: Social Science Center.
[3] Eurostat (2018). *Regional statistics* [online]. Available at: http://ec.europa.eu/eurostat/web/regions-and-cities
[4] Freudenberg. M. (2003). Composite indicators of country performance: a critical assessment. *OECD STI Working Paper 2003/16*. pp. 1-34. Available at: http://www.oecd.org/sti/working-papers.
[5] Ginevičius. R. & Podvezko. V. (2005). Objective and subjective approaches in determining the criterion weights in multicriteria models. *Transport and telecommunication*, 6 (1), 133–137.
[6] Melecký, L. (2015). Assessment of Socioeconomic Development of Visegrad Four NUTS 2 Regions Using Composite Indices. In *Proceedings of 12th International Conference Economic Policy in the European Union Member Countries* (pp. 561–571). Karviná: Silesian University.
[7] Minarčíková. E. (2016). Application of selected weighting methods and TOPSIS method in regional disparities analysis. In *10th International Days of Statistics and Economics Conference Proceedings* (pp. 1240–1249). Prague: Melandrium.
[8] Munda. G. & Nardo. M. (2005). *Constructing Consistent Composite Indicators: the Issue of Weights*. Luxembourg: Office for Official Publications of the European Communities.
[9] Nasierowski. W. (2016). Composite indexes economic and social performance: Do they provide valuable information? *Foundations of Management*, 8 (1), 167–174
[10] OECD (2008). *Handbook on Constructing Composite Indicators. Methodology and User Guide*. France: OECD.
[11] Romer. P. (1990). Endogenous technological change. *Journal of Political Economy*, 98 (5), S71-S102.
[12] Saisana. M. & Tarantola. S. (2002). *Knowledge Economy Indicators. State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development*. Italy: European Communities.
[13] Saisana. M.. Tarantola. S.. Schulze. N.. Laurens. Ch.. Moesen. W. & Puyenbroeck. T.V. (2005). Knowledge Economy Indicators. State-of-the-Art Report on Composite Indicators for the Knowledge-based Economy *Workpackage 5 - CIS8–CT–2004–502529 KEI*.
[14] Sala-i-Martin, X. (2006). The world distribution of income: Falling poverty and … Convergence. Period. *The Quarterly Journal of Economics*, 121 (2), 351–397.
[15] Sharpe. A. & Andrews. B. (2012). An Assessment of Weighting Methodologies for Composite Indicators: The Case of the Index of Economic Well-being. *CSLS Research Report. No. 2012-10*.

On consistency and coverage measures in the fuzzified set-theoretic approach for social sciences: dealing with ambivalent evidence in the data

Jan Stoklasa[1], Tomáš Talášek[2], Pasi Luukka [3]

Abstract. This paper investigates the modified concepts of fuzzy consistency and fuzzy coverage recently proposed by Stoklasa et al. in the context of the fuzzified set-theoretic approach for the purpose of linguistic/fuzzy rule validation in social sciences. The new fuzzified consistency and coverage measures (denoted by the F_2 and F_3 indices in the paper by Stoklasa et al.) were introduced in 2017 as a reaction to the inability of the original consistency and coverage fuzzified by Ragin in 2008 to deal with ambivalent information present in the data. More specifically, the F_2 and F_3 measures try to remove the ambivalent information inherent in the fuzzy observations - i.e. observations with fuzzy values of some or all of the relevant variables. We analyse the performance of all three consistency and coverage measures, assess the ability of the new ones to deal with the ambivalent information and propose a modification of the F_3 consistency and coverage which increases its distinguishing power. We present graphical summaries of the performance of the consistency and coverage measures for a single-fuzzy-observation case and discuss their possible usefulness and differences.

Keywords: Set-theoretic, consistency, coverage, fuzzification, ambivalence.

JEL classification: B40, C44
AMS classification: 97M40

1 Introduction

The set-theoretic approach proposed by Ragin [6] in the field of political sciences and further popularized by the same author in social sciences in general [7, 8], and by Fiss [3] also in the context of organizational research, provides easy-to-use tools for the assessment of set relations in the data. These set relations can be considered to represent patterns in the data that can reflect verbally described relationships within the studied system. As such the set-theoretic approach offers to social sciences and humanities much needed tools for the validation of verbally formulated assumptions using available data. The possible applications of this approach in economics cover almost all areas of interest ranging from organizational configurations research [3] through investment analysis [2] and poverty research [5] to tourism success modeling [10].

Deep within the set-theoretic methodology lie the concepts of *consistency* (in fact a measure of ambivalence or contradiction within the data set or truth table - in other words measure of support of the given relationship between the given conditions and the outcome in the data) and *coverage* (a measure of necessity of the conditions for the given outcome, as reflected in the data set). It is these concepts, or to be more specific their fuzzifications, that are directly in the scope of this paper.

Since the introduction of fuzzy sets and linguistic variables [17, 18], fuzzy sets and the linguistic modelling framework [14] have been considered a potentially useful tool for social sciences and humanities (see e.g. [1, 12, 13, 16, 19]) . The imprecision inherent in social science data requires appropriate tools for its reflection. It is thus of no surprise that the crisp set-theoretic concepts and methods were fuzzified [4, 8, 9] to enable the use of these methods also in connection with linguistic descriptions of the studied systems. The appeal of the set-theoretic approach in its crisp and fuzzified version lies in its simplicity (the use of set relations) and in the ability to provide conclusions even for small data sets. It can also be used for theory building purposes within the fuzzy set Qualitative Comparative Analysis (fsQCA, [11, 8]) - i.e. for the assessment of the validity of assumptions and claims (represented e.g. by (fuzzy) rules) and their modifications which are most in line with the data gathered from the studied system. Even though the fuzzified version of the set-theoretic approach and the fsQCA using

[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

[3]Lappeenranta University of Technology, School of Business and Management, Skinnarilankatu 34, Lappeenranta, Finland, pasi.luukka@lut.fi

the fuzzified consistency and coverage concepts (as introduced by Ragin in [8]) has already made its way to university textbooks, there are still some doubts concerning the appropriateness of the fuzzification of the crisp set-theoretic concepts. Stoklasa, Luukka and Talášek [15] point out that the direct fuzzification of the crisp set-theoretic concepts of consistency and coverage as proposed in [8] is possible, but introduces severe complications for the interpretation of the values of these fuzzified measures. Since in the fuzzy context, the same data instance can provide partial evidence in favour of the given rule, and at the same time partially provide counterevidence to the very same rule, a redefinition of these measures for the fuzzy-set-theoretic analysis might be more appropriate than their direct fuzzification. In this paper, we aim to shed more light on this claim, analyze the performance of Ragin's fuzzy set consistency and coverage and the redefined measures of consistency and coverage proposed by Stoklasa et al. We also suggest a modified consistency and coverage measure for the fuzzified set-theoretic approach, that offers more interpretation possibilities than the studied measures presented in [8] and [15]. We also note, that the novel concepts of degree-of-support and degree-of-disproof introduced in [15] are left out of the scope of this paper. The rest of the paper is organized as follows: first we recall some needed definitions of fuzzy-set concept and the definitions of the set-theoretic consistency and coverage, their fuzzified versions by Ragin [8] and Stoklasa et al. (the F_2 and F_3 consistency and coverage measures suggested in [15]) and their connection to rule validation. Then we analyze the performance on these measures on a simple case of a fuzzy rule validation by a single data instance and discuss their strengths and possible weaknesses (to maintain the simplicity of interpretation and graphical representation). Finally we suggest a modified version of the F_3 consistency and coverage and compare its performance with the previously suggested measures. We end the paper with a short discussion of the results.

2 Set-theoretic consistency and coverage and their fuzzifications

Let us consider a set of observations $U = \{x_1, x_2, \ldots, x_n\}$. Let us consider a feature A and an indicator function $\chi^A : U \to \{0, 1\}$ such that $\chi^A(x_i) = 1$ if and only if x_i has the feature A and $\chi^A(x_i) = 0$ otherwise, for all $i = 1, \ldots, n$. Let us also consider a feature B with an analogous indicator function $\chi^B : U \to \{0, 1\}$. Let us also introduce a negation of the feature B representing the absence of the feature B (denoted B' and meaning "not B"), for which the indicator function is $\chi^{B'} : U \to \{0, 1\}$ such that $\chi^{B'}(x_i) = 1$ if and only if x_i does not have the feature B and $\chi^{B'}(x_i) = 0$ otherwise. Now let us consider that we would like to investigate the assumption that *IF an observation has a feature A, THEN it has the feature B as well*, or $A \Rightarrow B$ for short. Given the set of observations U, we can assess the consistency [8] of such an assumption with the data (its support by the data) computing the consistency of $A \Rightarrow B$:

$$\text{Consistency}(A \Rightarrow B) = \frac{\sum_{i=1}^{n} \min\{\chi^A(x_i), \chi^B(x_i)\}}{\sum_{i=1}^{n} \chi^A(x_i)} = \frac{\text{Card}(A \cap B)}{\text{Card}(A)}, \tag{1}$$

where Card(A) represents the cardinality of the set A, i.e. the number of its elements, and $\cap$ is the standard set intersection, i.e. $\chi^{(A \cap B)}(x_i) = \min\{\chi^A(x_i), \chi^B(x_i)\}$. Note, that U is fully consistent with $A \Rightarrow B$ as long as $A \subseteq B$, i.e. in this case Consistency$(A \Rightarrow B) = 1$ and we can interpret this as the absence of counterexamples to $(A \Rightarrow B)$; obviously we assume that Card$(A) \neq 0$. Analogously we can calculate a measure of "universality" of the assumption $A \Rightarrow B$ for the given set of observations U as the coverage of $A \Rightarrow B$ (assuming that Card$(B) \neq 0$):

$$\text{Coverage}(A \Rightarrow B) = \frac{\sum_{i=1}^{n} \min\{\chi^A(x_i), \chi^B(x_i)\}}{\sum_{i=1}^{n} \chi^B(x_i)} = \frac{\text{Card}(A \cap B)}{\text{Card}(B)}. \tag{2}$$

We can clearly see that Coverage$(A \Rightarrow B) = 1$ if and only if $B \subseteq A$. In other words both measures are based on subsethood. These measures and their use is without problems as long as we consider A and B to be crisp (non-fuzzy) sets, i.e. as long as the characteristic features of A and/or B are either present or not present in each observation.

Since the set-theoretic approach was introduced to social sciences (see [6, 7]), its fuzzification was introduced by Ragin in [8] and finally adopted in the fsQCA framework. Let us therefore from now on assume that A and B are fuzzy sets and $\mu_A : U \to [0, 1]$ and $\mu_B : U \to [0, 1]$ are their respective membership functions. The intersection of two fuzzy sets A and B on the same universe U is a fuzzy set $(A \cap B)$ on U with the membership function $\mu_{A \cap B} : U \to [0, 1]$ such that for any $x \in U$ we have $\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\}$. A is a fuzzy subset of B (denoted $A \subseteq_F B$) if for all $x \in U$ it holds that $\mu_A(x) \leq \mu_B(x)$. Ragin's direct fuzzification of (1) and (2) stemming from the subsethood interpretation of consistency and coverage is expressed by the following formulas (we will denote Ragin's versions of the fuzzified formulas by the subscript F):

$$\text{Consistency}_F(A \Rightarrow B) = \frac{\sum_{i=1}^{n} \min\left(\mu_A(x_i), \mu_B(x_i)\right)}{\sum_{i=1}^{n} \mu_A(x_i)}, \tag{3}$$

$$\text{Coverage}_F(A \Rightarrow B) = \frac{\sum_{i=1}^{n} \min\left(\mu_A(x_i), \mu_B(x_i)\right)}{\sum_{i=1}^{n} \mu_B(x_i)}. \tag{4}$$

As pointed out by Stoklasa et al. [15], the direct fuzzification embodied by (3) and (4) dramatically changes the interpretation of the consistency and coverage. Many authors suggest the use of XY plots to assess the $A \subseteq_F B$ fuzzy subset relation to assess e.g. the sufficiency of A for B (see e.g [8, 11]). These authors consider $A \subseteq_F B$ to be a strong enough evidence of the fuzzy subset relation being consistent with the sufficiency of A for B. The main drawback of this approach, as pointed out in [15], is that as long as A and B are fuzzy sets, the same data can provide partial evidence in favor of $A \Rightarrow B$ and $A \Rightarrow B'$ at the same time (see Figure 1).

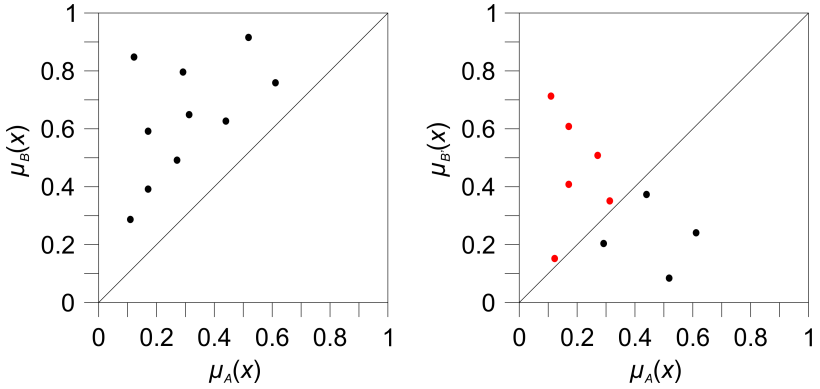


Figure 1 An example of a data set ($n = 10$) fully consistent with A being a sufficient condition for B according to (3) - left subfigure. In the right subfigure the same data set is used to assess the sufficiency of A for B'; red data-points support both $A \Rightarrow B$ and $A \Rightarrow B'$ at the same time.

Since consistency is linked with the investigation of the $A \subseteq B$ subset relation (or $A \subseteq_F B$ relation in the fuzzy case) and coverage investigates analogously the opposite relationship $B \subseteq A$ (or $B \subseteq_F A$ in the fuzzy case), we will focus on the investigation of consistency from now on without any loss of generality. Notice, that in the case of (3) it is possible to obtain full confirmation of $A \Rightarrow B$ and $A \Rightarrow B'$ at the same time - as pointed out by Stoklasa et al. [15]. The reason for this is, that as long as A and B are fuzzy sets, we can have $A \cap B \neq \emptyset$ and $A \cap B' \neq \emptyset$ at the same time. In other words the same data can provide evidence for $A \Rightarrow B$ and also disproof this very relationship (providing evidence for $A \Rightarrow B'$) at the same time. The use of (3) and (4) is therefore constitutes a high risk of misinterpretation of the information in the data. To see this issue more clearly, let us now consider a single data point x and let us plot how the variation of its membership degree to the fuzzy sets A and B influences the consistency of x with $A \Rightarrow B$ and also the consistency of x with $A \Rightarrow B'$. Figure 2 summarizes the situation - yellow points represent such data-points x for which their membership degree to the fuzzy sets A and B (the coordinates of the points in the subfigures) result in full consistency with the respective investigated relation. We can clearly see that there is a nonempty intersection of the yellow areas in the left and right subfigures, i.e. there are many instances of x which are fully consistent with $A \Rightarrow B$ and $A \Rightarrow B'$. These observations should be considered completely *ambivalent*, since they provide full evidence for the investigated relationship and at the same time full evidence against it. It therefore seems reasonable either to always investigate the consistency of the data with $A \Rightarrow B$ and $A \Rightarrow B'$ at the same time, but never separately. Alternatively, measures of consistency that remove the effect of the ambivalent observations can be defined. Stoklasa et al. [15] suggest the formulas (5) and (6) to deal with the issue of ambivalent observations (assuming $\sum_{i=1}^{n} \mu_A(x_i) \neq 0$ in both of them). In essence (5) removes from the support of the given relationship represented by $A \subseteq B$ that part that corresponds with $A \subseteq (B \cap B')$, i.e. that piece of information that simultaneously supports both the outcome B and its negation B' under the condition A. On the other hand (6) balances the amount of evidence in favour of $A \Rightarrow B$ and in favour of $A \Rightarrow B'$, the amount of evidence in favour of $A \Rightarrow B$ is reduced by the counter-evidence to this relationship (i.e. by the amount of evidence in favour of $A \Rightarrow B'$); if the counter-evidence is stronger than the evidence, then $\text{Consistency}_{F_3}(A \Rightarrow B) = 0$.

$$\text{Consistency}_{F_2}(A \Rightarrow B) = \frac{\sum_{i=1}^{n} (\min\left(\mu_A(x_i), \mu_B(x_i)\right) - \min\left(\mu_A(x_i), \mu_B(x_i), \mu_{B'}(x_i)\right))}{\sum_{i=1}^{n} \mu_A(x_i)} \tag{5}$$

$$\text{Consistency}_{F_3}(A \Rightarrow B) = \max\left\{0; \frac{\sum_{i=1}^{n} (\min\left(\mu_A(x_i), \mu_B(x_i)\right) - \min\left(\mu_A(x_i), \mu_{B'}(x_i)\right))}{\sum_{i=1}^{n} \mu_A(x_i)}\right\} \tag{6}$$

Theorem 1. *For a single observation case formulas* (5) *and* (6) *coincide.*

Proof. Let us consider a single observation x and a relationship $A \Rightarrow B$ to be investigated.

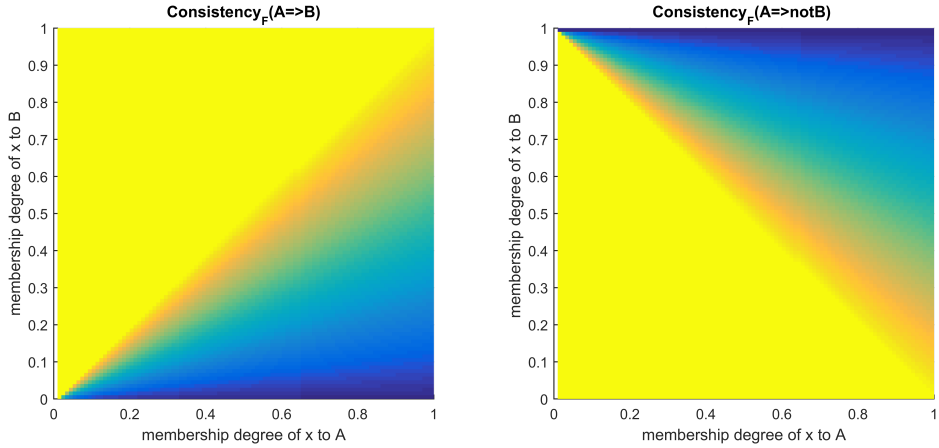


Figure 2 The consistency of the observation x with $A \Rightarrow B$ (left subfigure) and the consistency of x with $A \Rightarrow B'$ (right subfigure); consistency is computed by (3). Yellow colour represents full consistency (i.e. the consistency of 1), dark blue colour represents the consistency of 0.

- Let us first consider $\mu_A(x) \geq \mu_B(x) \geq 0.5$. From (5) we have $\text{Consistency}_{F_2}(A \Rightarrow B) = \frac{\mu_B(x) - \mu_{B'}(x)}{\mu_A(x)}$, from (6) we have $\text{Consistency}_{F_3}(A \Rightarrow B) = \max\{0; \frac{\mu_B(x) - \mu_{B'}(x)}{\mu_A(x)}\} = \frac{\mu_B(x) - \mu_{B'}(x)}{\mu_A(x)} = \text{Consistency}_{F_2}(A \Rightarrow B)$.
- For $\mu_A(x) \geq \mu_B(x)$ and $\mu_B < 0.5$ we get $\text{Consistency}_{F_2}(A \Rightarrow B) = \frac{\mu_B(x) - \mu_B(x)}{\mu_A(x)} = 0$ and $\text{Consistency}_{F_3}(A \Rightarrow B) = \max\{0; \frac{\mu_B(x) - \min\{\mu_{B'}(x); \mu_A(x)\}}{\mu_A(x)}\} = 0$ since $\mu_{B'}(x) > \mu_B(x)$ and $\mu_A(x) \geq \mu_B(x)$.
  Hence $\text{Consistency}_{F_2}(A \Rightarrow B) = \text{Consistency}_{F_3}(A \Rightarrow B)$.
- For $\mu_A(x) < \mu_B(x) < 0.5$ we get $\text{Consistency}_{F_2}(A \Rightarrow B) = \frac{\mu_A(x) - \mu_A(x)}{\mu_A(x)} = 0$ and $\text{Consistency}_{F_3}(A \Rightarrow B) = \max\{0; \frac{\mu_A(x) - \mu_A(x)}{\mu_A(x)}\} = 0$. Again $\text{Consistency}_{F_2}(A \Rightarrow B) = \text{Consistency}_{F_3}(A \Rightarrow B)$.
- Finally for $\mu_A(x) < \mu_B(x)$ and $\mu_B(x) \geq 0.5$ we get $\text{Consistency}_{F_2}(A \Rightarrow B) = \frac{\mu_A(x) - \min\{\mu_{B'}(x); \mu_A(x)\}}{\mu_A(x)}$ and $\text{Consistency}_{F_3}(A \Rightarrow B) = \max\{0; \frac{\mu_A(x) - \min\{\mu_{B'}(x); \mu_A(x)\}}{\mu_A(x)}\} = \frac{\mu_A(x) - \min\{\mu_{B'}(x); \mu_A(x)\}}{\mu_A(x)}\}$.
  Again $\text{Consistency}_{F_2}(A \Rightarrow B) = \text{Consistency}_{F_3}(A \Rightarrow B)$.

□

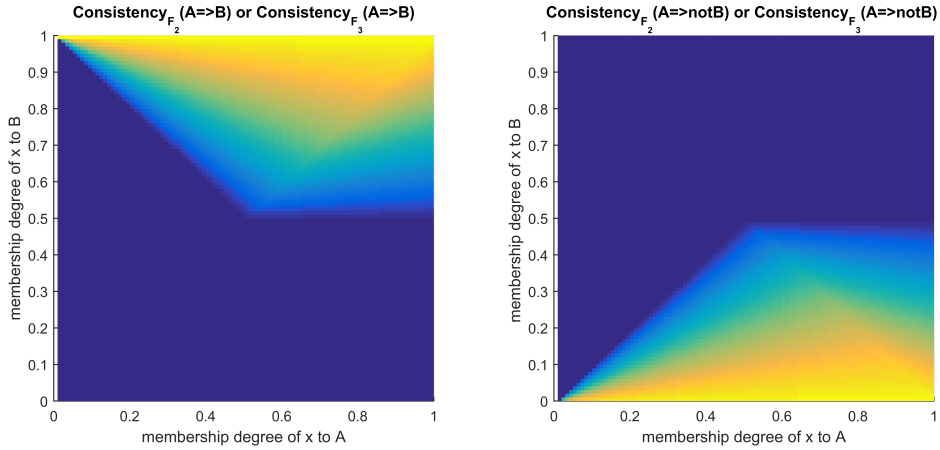


Figure 3 The consistency of the observation x with $A \Rightarrow B$ (left subfigure) and the consistency of x with $A \Rightarrow B'$ (right subfigure). consistency is computed by (5) or (6) which both give the same values for the single-observation case. Yellow colour represents full consistency (i.e. the consistency of 1), dark blue colour represents the consistency of 0.

Based on the above-proven theorem the F_2 and F_3 consistency measures give the same consistency values for the single-observation case. We can therefore analyze their performance simultaneously. Figure 3 again assumes the single-observation perspective and plots full consistency of the observation in yellow colour and zero consis-

tency in dark-blue (coordinates represent the membership degrees of the observation x to A and B respectively). It clearly shows that there is no overlap of yellow areas any more between the left and right subplot. In other words there are no more observations that would provide simultaneous support and disproof w.r.t. the given relationship. The F_2 and F_3 consistency measures thus perform as expected and as intended by Stoklasa et al. [15], i.e. they do not consider the ambivalent observations to provide any support for the investigated relationship.

3 A modified fuzzy consistency and coverage measure

Although (5) and (6) provide means for the removal of the influence of observations representing ambivalent evidence in favour of and against the given relationship. Some information is lost when we use these consistency measures. Particularly these measures do not distinguish between observations providing ambivalent evidence and observations providing no evidence in favour and against the given relationship. We can, however, modify the formula (6) in such a way that would distinguish between ambivalent evidence and no evidence. We therefore suggest a new consistency measure F_4 defined by (7), again assuming that $\sum_{i=1}^{n} \mu_A(x_i) \neq 0$. Using this measure, observations that carry ambivalent information (support $A \Rightarrow B$ and $A \Rightarrow B'$ at the same time) result in 0.5 consistency (in the single-observation case), while observations not providing any support for $A \Rightarrow B$ result in a zero consistency with $A \Rightarrow B$. Figure 4 summarizes the performance of this new consistency measure.

$$\text{Consistency}_{F_4}(A \Rightarrow B) = \frac{1}{2}\left(1 + \frac{\sum_{i=1}^{n}(\min(\mu_A(x_i), \mu_B(x_i)) - \min(\mu_A(x_i), \mu_{B'}(x_i)))}{\sum_{i=1}^{n} \mu_A(x_i)}\right) \tag{7}$$

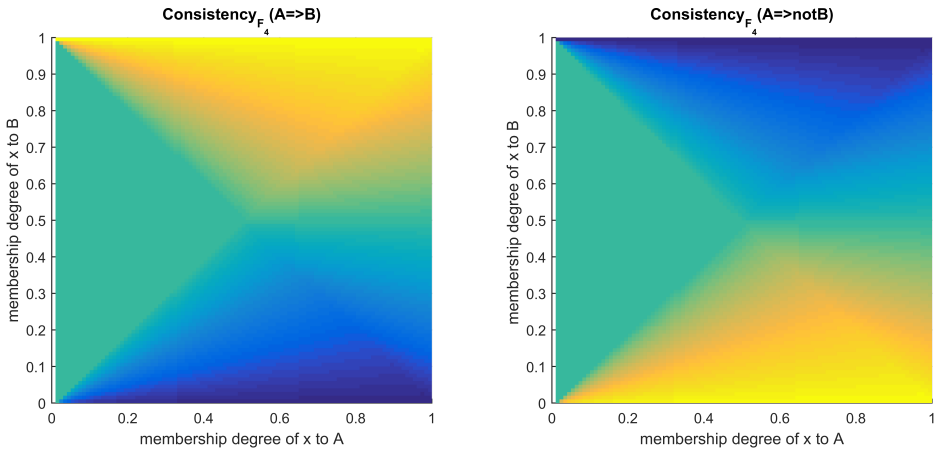


Figure 4 The consistency of the observation x with $A \Rightarrow B$ (left subfigure) and the consistency of x with $A \Rightarrow B'$ (right subfigure). consistency is computed by the newly proposed consistency formula (7). Yellow colour represents full consistency (i.e. the consistency of 1), dark blue colour represents the consistency of 0, shades of green represent intermediate consistency values.

Not only does this newly proposed measure distinguish between the ambivalent observations and observations containing no relevant information for the investigation of the given relationship, it also satisfies the following property represented by Theorem 2. We can see, that the consistency measure F_4 distributes the support in the data between $A \Rightarrow B$ and $A \Rightarrow B'$. It can therefore be used as an indicator whether the data seem to support more $A \Rightarrow B$ or rather $A \Rightarrow B'$. Note, that the F_4 consistency of 0.5 signifies the inability to decide whether B or B' is the outcome of A with respect to the data. Stoklasa et al. [15] suggest the degree-of-support and degree-of-disproof measures to investigate the support for the given relationship (and disproof thereof) in more details on various levels of (un)certainty.

Theorem 2. *For any set of observations* $\{x_1, \ldots, x_n\}$ *the following holds:*

$$\textit{Consistency}_{F_4}(A \Rightarrow B) + \textit{Consistency}_{F_4}(A \Rightarrow B') = 1. \tag{8}$$

Proof. Calculating the left hand side of (8) we get:
$\frac{1}{2}\left(1 + \frac{\sum_{i=1}^{n}(\min(\mu_A(x_i),\mu_B(x_i)) - \min(\mu_A(x_i),\mu_{B'}(x_i)))}{\sum_{i=1}^{n}\mu_A(x_i)}\right) + \frac{1}{2}\left(1 + \frac{\sum_{i=1}^{n}(\min(\mu_A(x_i),\mu_{B'}(x_i)) - \min(\mu_A(x_i),\mu_B(x_i)))}{\sum_{i=1}^{n}\mu_A(x_i)}\right) =$
$1 + \frac{\sum_{i=1}^{n}(\min(\mu_A(x_i),\mu_B(x_i)) - \min(\mu_A(x_i),\mu_{B'}(x_i)) + \min(\mu_A(x_i),\mu_{B'}(x_i)) - \min(\mu_A(x_i),\mu_B(x_i)))}{2\sum_{i=1}^{n}\mu_A(x_i)} = 1$ □

4 Conclusions

This paper studies the notions of consistency and coverage as proposed by Ragin [6] for the set-theoretic approach and their fuzzifications proposed in [8, 15]. We suggest a visualization of the performance of the consistency measures analogous to the XY plots used in set-theoretic and fsQCA textbooks and confirm the reasonability of the modifications of fuzzified set-theoretic concepts of consistency and coverage proposed in [15]. We prove that for a single observation the F_2 and F_3 return the same values. We show that the intention of their creators to weaken the influence of observations providing ambivalent evidence in favour and against the investigated relationship at the same time is working well. We also propose a modification of the F_3 consistency that distinguishes between the ambivalent evidence and missing evidence in the data. The newly proposed consistency F_4 also distributes the evidence available in the data completely between the support of $A \Rightarrow B$ and $A \Rightarrow B'$. We have stressed the importance of proper fuzzification of the non-fuzzy set-theoretic measures of consistency and coverage as pointed out already in [15] and contribute to the toolbox of fsQCA by proposing a new approach to consistency assessment.

Acknowledgements

Supported by the grant No. IGA_FF_2018_002 of the Internal Grant Agency of Palacký University Olomouc.

References

[1] Arfi, B.: *Linguistic fuzzy logic methods in social sciences*. Springer-Verlag, Berlin Heidelberg, 2010.

[2] Fan, D., Cui, L., Li, Y., and Zhu, C. J.: Localized learning by emerging multinational enterprises in developed host countries: A fuzzy-set analysis of Chinese foreign direct investment in Australia. *International Business Review* **25** (2016), 187–203.

[3] Fiss, P. C.: A Set-Theoretical Approach to Organizational Configurations. *Academy of Management Review* **32** (2007), 1180–1198.

[4] Fiss, P. C.: Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal* **54** (2011), 393–420.

[5] Neff, D.: Fuzzy set theoretic applications in poverty research. *Policy and Society* **32** (2013), 319–331.

[6] Ragin, C. C.: *The comparative method: Moving beyond qualitative and quantitative strategies*. University of California Press, Berkeley, Los Angeles, London, 1989.

[7] Ragin, C. C.: Set relations in social research: Evaluating their consistency and coverage. *Political Analysis* **14** (2006), 291–310.

[8] Ragin, C. C.: *Redesigning Social Inquiry: Fuzzy sets and Beyond*. University of Chicago Press, Chicago, 2008.

[9] Ragin, C. C.: *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press, Oakland, 2014.

[10] Rey-Martí, A., Ribeiro-Soriano, D., and Palacios-Marqués, D.: Entrepreneurial attributes of human capital and contingency factors in the culinary tourism. *International Entrepreneurship and Management Journal* **12** (2016), 67–85.

[11] Schneider, C. Q., and Wagemann, C.: *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge University Press, Cambridge, 2012.

[12] Smithson, M.: Fuzzy Sets and Fuzzy Logic in the Human Sciences. In: *Fuzzy Logic in Its 50th Year* (Kahraman, C., Kaymak, U., and Yazici, A., eds.), volume 341. Springer international publishing, 2016, 175–186.

[13] Smithson, M., and Verkuilen, J.: *Fuzzy set theory: applications in the social sciences*. Sage Publications, Thousand Oaks, London, New Delhi, 2006.

[14] Stoklasa, J.: *Linguistic models for decision support*. Lappeenranta University of Technology, Lappeenranta, 2014.

[15] Stoklasa, J., Luukka, P., and Talášek, T.: Set-theoretic methodology using fuzzy sets in rule extraction and validation - consistency and coverage revisited. *Information Sciences* **412-413** (2017), 154–173.

[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] Zadeh, L. A.: Fuzzy sets. *Information and control* **8** (1965), 338–353.

[18] Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences* **8** (1975), 199–249.

[19] Zétényi, T., ed.: *Fuzzy sets in psychology*. North-Holland, Amsterdam, New York, Oxford, Tokyo, 1988.

Fuzzy consensus in group decision-making model using absolute-type evaluations

Vojtěch Sukač[1], Ondřej Pavlačka[2]

Abstract. In the paper, we consider the fuzzy group decision-making model, which has been introduced in recent years. An absolute-type evaluation of alternatives is used in this model. This evaluation express the degree of fulfillment of the given goal by the respective alternative.
This model is based on the idea that an optimal alternative is chosen from the subset of alternatives, which are sufficiently good (user-defined) according to enough (user-defined) important (user-defined) experts. In the other words, we drop out alternatives which are evaluated very well by some important experts but according to other important experts are evaluated bad. On the contrary, we introduce a different approach to determine the strength of the consensus than in the original model. It turned out that the new approach can form a different subset of sufficiently consensual alternatives. This fact is illustrated by an example.

Keywords: Fuzzy, group decision-making, absolute-type evaluation, fuzzy consensus.

JEL classification: C44
AMS classification: 90B50

1 Introduction

In situations where a group of more people are deciding about the same issue, it is usually difficult to find a solution that everyone would be comfortable with, as everyone generally has different preferences. For these situations, a lot of group decision-making models can be found in the literature. Generally, they can be divided into two groups; the voting systems on the one hand and consensus reaching models on the other.

In this paper we recall and analyse the fuzzy group decision-making model proposed in [5, 6, 7]. Consensus reaching in the model analyzed in this paper is based on the idea that the choice of the best alternative should be made only among the alternatives that are good enough according to most of the relevant experts. Such preselection of alternatives is based on the idea of 'soft' consensus introduced by Kacprzyk and Fedrizzi [1, 2]. The model utilizes fuzzy evaluations of absolute type. These evaluations can be interpreted as the uncertain degrees of fulfilment of given goals [8]. Such evaluations are desirable, since they are not dependent on the set of alternatives and describe the acceptability of the alternatives (this also removes the *"one-eyed is king among the blind" effect* from the evaluation).

The above mentioned model has not good properties in case where the fuzzy evaluations of alternatives are too vague; we will illustrate this by an example. Therefore the aim of the paper is to present a modified approach that deals with the uncertainty of fuzzy variables more appropriatly. The modification will consist in different determination of the strength of the consensus. To the best of our knowledge, this modification has never been proposed in the literature so far.

The paper is organized as follows. The next section summarizes the concepts of fuzzy sets theory necessary to introduce the model. In the third section the original decision-making model is described in brief. In the fourth section the new approach is proposed. In the fifth section the results obtained from new approach are compared with the results of original model.

2 Preliminaries

Let U be a nonempty set called *universe*. *The fuzzy set* A on U is determined by its *membership function* $\mu_A(x) : U \to [0,1]$, where $\mu_A(x)$ expresses the *degree of membership* of x in the fuzzy set A — from 0 for "x definitely does not belong to A" to 1 for "x definitely belongs to A", through all the intermediate values. The family of all fuzzy sets on the universe U is denoted by $\mathcal{F}(U)$. Let A be a fuzzy set on U and $\alpha \in [0,1]$. The crisp set $A_\alpha = \{x \in U \,|\, \mu_A(x) \geq \alpha\}$ is called *the α-cut of the fuzzy set A*. *The support of the fuzzy set A* is a (crisp) set $\mathrm{Supp}(A) = \{x \in U \,|\, \mu_A(x) > 0\}$. *The kernel of the fuzzy set A* is a (crisp) set $\mathrm{Ker}(A) = \{x \in U \,|\, \mu_A(x) = 1\}$.

[1]Palacký University Olomouc, Department of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/12, 77146 Olomouc, Czech Republic, vojta.sukac@gmail.com
[2]dtto, ondrej.pavlacka@upol.cz

In cases, when the support of A is a discrete set ($\mathrm{Supp}(A) = \{x_1, \ldots, x_k\}$), then the fuzzy set A can be denoted as $A = \{{}^{\mu_A(x_1)}\!/_{x_1}, \ldots, {}^{\mu_A(x_k)}\!/_{x_k}\}$. A special type of fuzzy sets whose universe is a subset of $\mathbb{R}$, the so called *fuzzy numbers*, can be defined in the following way. Let $U \subset \mathbb{R}$ be an interval. *The fuzzy number* C is a fuzzy set on U which fulfills the following conditions:

a) $\mathrm{Ker}(C) \neq \emptyset$;
b) for all $\alpha \in (0, 1]$, C_α are closed intervals;
c) $\mathrm{Supp}(C)$ is bounded.

The family of all fuzzy numbers on U is denoted by $\mathcal{F}_N(U)$.

Each fuzzy number C is determined by $C = \{[\underline{C}(\alpha), \overline{C}(\alpha)]\}_{\alpha\in[0,1]}$, where $\underline{C}(\alpha)$ and $\overline{C}(\alpha)$ is the lower and the upper bound of the α-cut of the fuzzy number C respectively, for $0 < \alpha \leq 1$ and $[\underline{C}(0), \overline{C}(0)]$ is the closure of the support of C, i.e. $[\underline{C}(0), \overline{C}(0)] = \mathrm{Cl}(\mathrm{Supp}(C))$. A trapezoidal fuzzy number C is determined by the ordered quadruple $(c_1, c_2, c_3, c_4) \in U^4$ of the significant values of C satisfying $[c_1, c_4] = \mathrm{Cl}(\mathrm{Supp}(C))$ and $[c_2, c_3] = \mathrm{Ker}(C)$. The membership function of a trapezoidal fuzzy number C is given by

$$\mu_C(x) = \begin{cases} \frac{x-c_1}{c_2-c_1} & \text{if } c_1 \leq x < c_2, \\ 1 & \text{if } c_2 \leq x \leq c_3, \\ \frac{c_4-x}{c_4-c_3} & \text{if } c_3 < x \leq c_4, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

A fuzzy number C is called a *triangular fuzzy number* if $c_2 = c_3$. Closed real intervals and real numbers can be represented by special cases of trapezoidal fuzzy numbers. See e.g. [3] for more details.

The paper utilizes the linguistic approach to group decision-making problems. *A linguistic variable* [10] is the 5-tuple $(\mathcal{V}, \mathcal{T}(\mathcal{V}), U, G, M)$, where $\mathcal{V}$ is the name of the linguistic variable, $\mathcal{T}(\mathcal{V})$ is the set of its linguistic terms (the values of $\mathcal{V}$), $U \subset \mathbb{R}$ is the universe on which fuzzy numbers expressing meanings of these linguistic terms are defined, G is grammar used for generating of linguistic terms $\mathcal{T}(\mathcal{V})$ and M is a mapping that assigns to each term $\mathcal{C} \in \mathcal{T}(\mathcal{V})$ its meaning $C = M(\mathcal{C})$ (a fuzzy number on U).

3 Description of the original model

For the purpose of this paper we will briefly introduce the original model. We consider the group decision-making problem addressed in [6] with a group $\boldsymbol{E} = \{E_1, \ldots, E_p\}$ of $p \geq 2$ individuals (experts) with possibly different competencies and a set $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ of $n \geq 2$ alternatives. Experts' competencies are given by trapezoidal fuzzy numbers L^k, $k = 1, \ldots, p$, on the interval $[0, 1]$, where the endpoints of this continuum have the following interpretations: 0 means an incompetent expert and 1 means a fully competent expert.

Each expert provides fuzzy evaluations of each alternative X_i. These fuzzy evaluations H_i^k, $i = 1, \ldots, n$, $k = 1, \ldots, p$, are provided in the form of trapezoidal fuzzy numbers on the interval $[0, 1]$ and express the degree of fulfilment of the given goal by an alternative X_i according to an expert E_k (0 is interpreted as the given goal is not fulfiled at all, 1 means a complete fulfilment of the given goal). These evaluations can also be the result of a multiple-criteria assessment of the given alternative by the given expert. As such these evaluations are of absolute type, that is they are not dependent on the set of alternatives and describe the acceptability of the alternatives for each expert.

Since the model is based on the idea that the optimal alternative should be chosen only from a subset of such alternatives in which a *sufficient quantity* of *important experts* agree that they are *good enough*, it has to be introduced some linguistic variables. The model can be summarized in the following steps (the exact procedure of defining the candidate set far exceeds the scope of this paper and its detailed description can be found in [6]):

1. First a linguistic variable $\widehat{\mathcal{A}}$ with the linguistic term set

$$\{\widehat{\mathcal{A}}_1, \ldots, \widehat{\mathcal{A}}_5\} = \{excellent, good, acceptable, borderline, unacceptable\}$$

   is introduced to express the level of acceptance of alternatives by experts. Using the values of this linguistic variable, a modified set of linguistic terms $\{\mathcal{A}_1, \ldots, \mathcal{A}_5\}$ is defined, where $\mathcal{A}_r$, $r = 1, \ldots, 5$, represents "$\widehat{\mathcal{A}}_r$ or better". The fuzzy number meanings $\widehat{A}_r$, A_r, $r = 1, \ldots, 5$, are expressed by fuzzy numbers on $[0, 1]$ (see figure 1).
2. The fuzzy sets F_{ri} of *experts suggesting* $\mathcal{A}_r$ *for* X_i, $F_{ri} \in \mathcal{F}(\{E_1, \ldots, E_p\})$, $r = 1, \ldots, 5,\ i = 1, \ldots, n$, are subsequently defined as $F_{ri} = \{{}^{\theta^1_{ri}}\!/_{E_1}, \ldots, {}^{\theta^p_{ri}}\!/_{E_p}\}$, where

$$\theta^k_{ri} = \sup_{x\in[0,1]} \big\{ \min\{\mu_{A_r}(x), \mu_{H_i^k}(x)\}\big\}, \quad k = 1, \ldots, p. \tag{2}$$

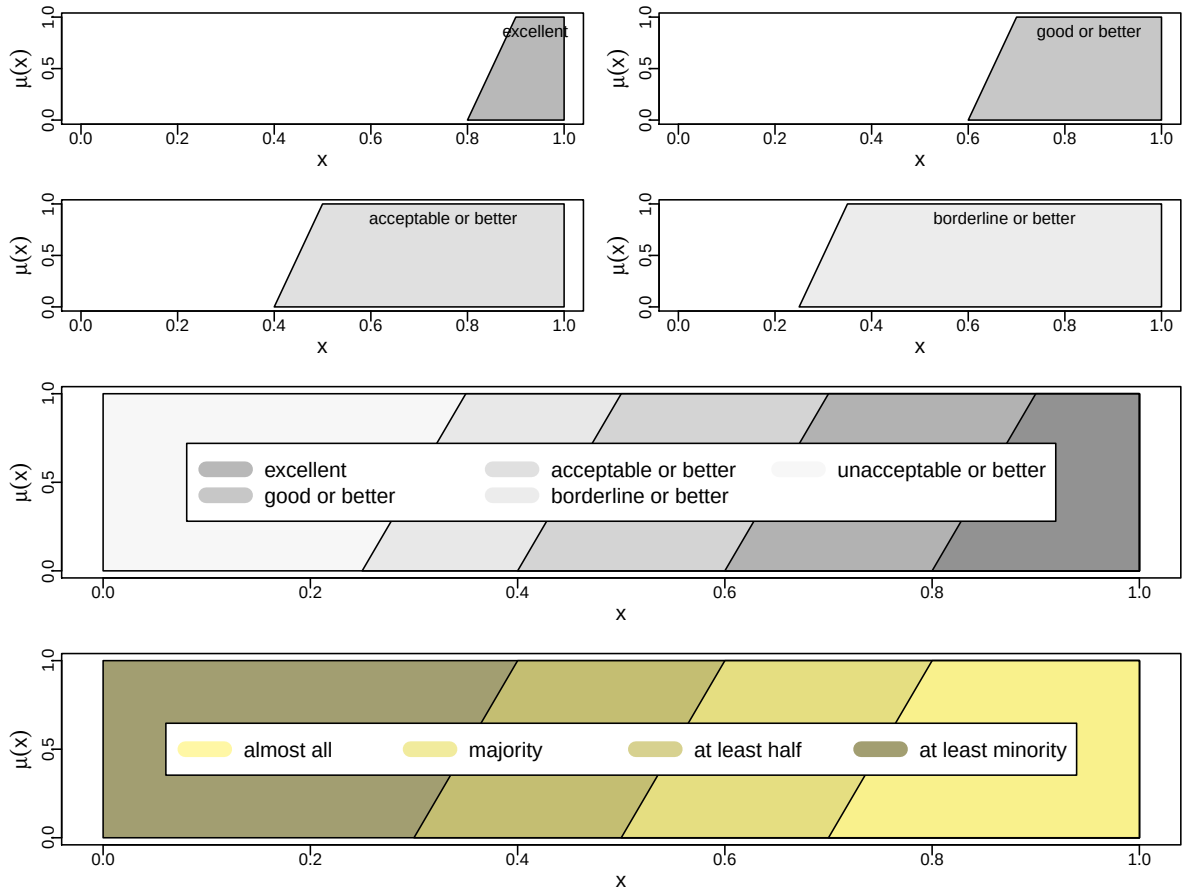


Figure 1 Linguistic terms $\mathcal{A}_r$ and $\mathcal{Q}_s$

3. The linguistic quantifier set

$$\{\widehat{\mathcal{Q}}_1, \ldots, \widehat{\mathcal{Q}}_4\} = \{almost\ all, more\ than\ half, about\ half, minority\}$$

is introduced and the meanings of the derived linguistic quantifiers $\{\mathcal{Q}_1, \ldots, \mathcal{Q}_4\}$, where $\mathcal{Q}_s$, $s = 1, \ldots, 4$, represents *"at least* $\widehat{\mathcal{Q}}_s$*"* are defined respectively by fuzzy numbers on $[0, 1]$ (see figure 1). The linguistic term *important expert* is also introduced, with its meaning $B \in \mathcal{F}_N([0, 1])$. The fuzzy set of *important experts* $I \in \mathcal{F}(\{E_1, \ldots, E_p\})$ is defined as $I = \{{}^{\zeta_1}\!/_{E_1}, \ldots, {}^{\zeta_p}\!/_{E_p}\}$, where

$$\zeta_k = \sup_{x \in [0,1]} \big\{ \min\{\mu_B(x), \mu_{L^k}(x)\} \big\}, \quad k = 1, \ldots, p. \tag{3}$$

4. The truth value of the statement "the alternative X_i is $\mathcal{A}_r$ (e.g. *at least good*) with respect to the opinion of quantity ($\mathcal{Q}_s$) of the important experts (I)" (denoted $\xi_i^{r,s}$) is determined, $i = 1, \ldots, n$, $r = 1, \ldots, 5$, $s = 1, \ldots, 4$ (see [6] for details).
5. For r, s, $r = 1, \ldots, 5$, $s = 1, \ldots, 4$, the sets $\Upsilon_{r,s}$ are defined, that include such alternatives that are $\mathcal{A}_r$ according to $\mathcal{Q}_s$ of the important experts, e.g.

$$\Upsilon_{r,s} = \{X_i \in \boldsymbol{X} \,\big|\, \xi_i^{r,s} = 1\}. \tag{4}$$

6. The order $\mathcal{K}$ is defined. The order $\mathcal{K}$ is a sequence of pairs $\big\{(r, s)_t\big\}_t$ of the values representing a given combination of r and s, such that the set $\Upsilon_{(r,s)_t}$ should be checked for nonemptyness before the set $\Upsilon_{(r,s)_{t'}}$, for $t < t'$. In our case we suggest to check first the set of alternatives considered (*excellent* by *almost all*) then the set of alternatives considered (*excellent* by *majority*), ..., and the last set we check for nonemptyness is the set of alternatives that are (*at least good* by *at least half*), e.g.

$$\mathcal{K} = \big\{(1, 1), (1, 2), \ldots, (2, 3)\big\}. \tag{5}$$

7. The *candidate set*, which is the first non-empty set $\Upsilon_{r,s}$ according to $\mathcal{K}$ (denoted Υ^*) includes those alternatives among which the most promising one is to be chosen.

Note that $\mathcal{K}$ can be defined to include only such alternatives that are still relevant for the given problem. E.g. looking for a set of alternatives that are at least borderline might not be appropriate if we are not willing to accept borderline alternatives. In this case pairs $(3, s)$ and $(4, s)$ will not be present in $\mathcal{K}$ for any value of s. Similarly we can reflect our requirements of a sufficient amount of experts that agree on the evaluation.

8. Finally, if the candidate set is found (i.e. there exists a non-empty set $\Upsilon_{r,s}$ according to the order $\mathcal{K}$), then any alternative from Υ^* could be chosen as the optimal one. One of the possible way to compare alternatives from Υ^* is to calculate the group evaluation of alternative using the fuzzy weighted average operation with the expert's competencies as fuzzy weights [4]. These group evaluations are then compared by the center of gravity method or any other suitable algorithm.

Note, that it can happen that there is no non-empty set $\Upsilon_{r,s}$ according to the order $\mathcal{K}$ (i.e. the candidate set is empty – or does not exists). In this case, no alternative is chosen (none of the alternatives is considered good enough – in the consensus sense – by the experts).

4 New approach

In the original model described in the previous section, the membership degree θ_{ri}^k is computed in the step 2 as the height of the intersection of the experts' fuzzy evaluation of alternative H_i^k and the meaning of the linguistic term $\mathcal{A}_r$ (2). As is shown in figure 2, by this procedure we completely lose the information about the uncertainty of the fuzzy evaluation (we will obtain only real number θ). Let us consider alternatives B and C with the same height of the intersection of their fuzzy evaluations and the meaning of the term *good or better*. It is desirable that a less uncertain alternative X_2 (the blue one) should be preferred to the more vague one X_1 (the red one). Due to this aim, we propose a modification of the method consisting in replacement of the height of the intersection by the relative size of the common area below the membership function of the intersection related to the area below the membership function of the fuzzy evaluation.

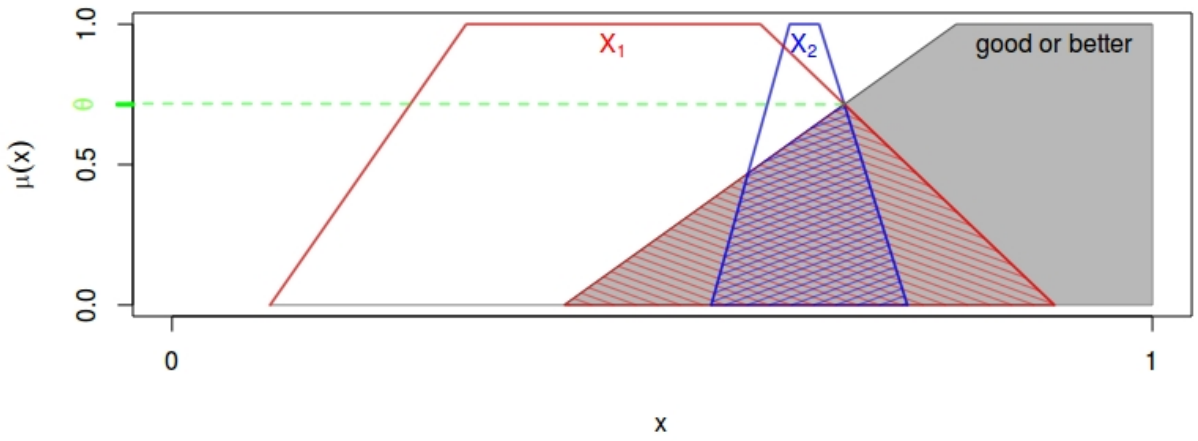


Figure 2 Comparing the acceptability levels computed by the original and the new approach

Note, that such modification is possible if and only if the fuzzy evaluation H_i^k is not a real number. If H_i^k is real, then the common area has a zero size. Therefore, if H_i^k is a real number, the formula (2) is left unchanged. Otherwise, the modified formula (2) has the form

$$\theta_{ri}^k = \frac{\int_0^1 \min\{\mu_{A_r}(x), \mu_{H_i^k}(x)\}\,\mathrm{d}x}{\int_0^1 \mu_{H_i^k}(x)\,\mathrm{d}x}, \quad i = 1, \ldots, n,\ r = 1, \ldots, 5,\ k = 1, \ldots, p. \tag{6}$$

The same modification is proposed to be applied in the step 3 with the height of the intersection of the experts' fuzzy competencies and the meaning of the linguistic term important expert (3). Similarly, this modification makes sense if experts competence L^k is not expressed by a real number. In this case, the modified formula (3) has the form

$$\zeta_k = \frac{\int_0^1 \min\{\mu_B(x), \mu_{L^k}(x)\}\,\mathrm{d}x}{\int_0^1 \mu_{L^k}(x)\,\mathrm{d}x}, \quad k = 1, \ldots, p. \tag{7}$$

The new approach in comparison with the original one deals with the whole membership functions of the fuzzy evaluations of alternatives and the fuzzy competencies of experts. Hence, it is clear that the modified approach at deals with the uncertainty of fuzzy variables more appropriately.

5 Example

Business management decides to buy a new production machine. There are four available alternatives on the market that meet the requirements. In order to achieve greater objectivity in decision-making, six experts were approached to provide the evaluation of each alternative. Their fuzzy evaluations that mean the fuzzy degrees of fulfilment of given goal are depicted in figure 3.

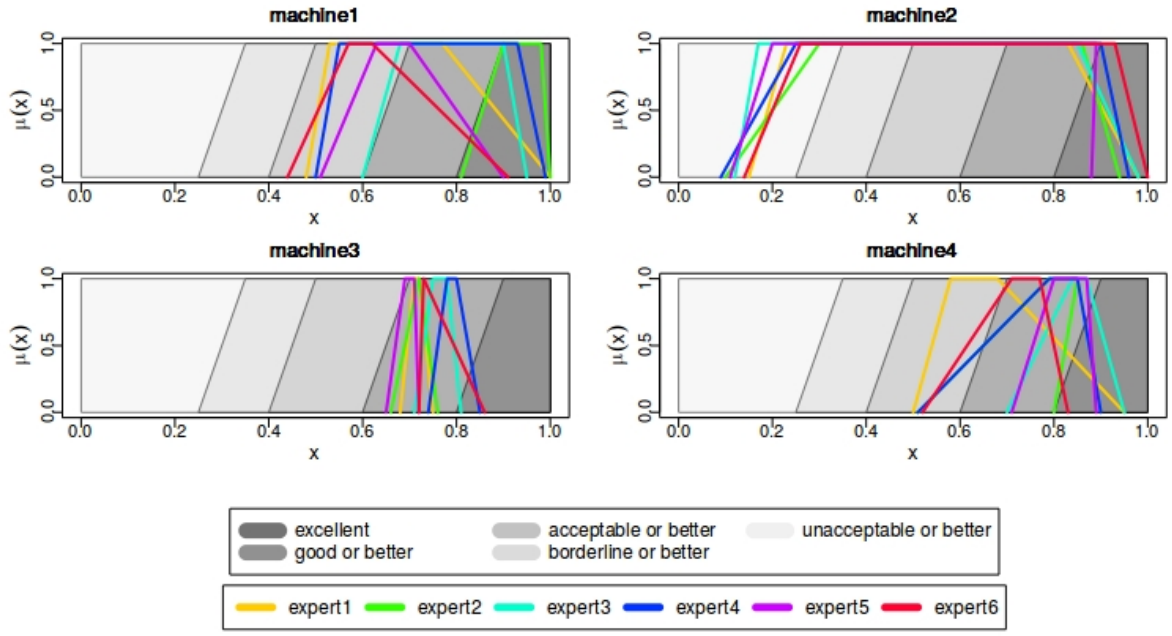


Figure 3 Expert's evaluation of alternatives

In accordance with the previous mentioned procedure, the candidate set is found. In table 1, we can see that according to the original approach, the candidate set contains only the alternavite *machine2* which is chosen as the optimal one. It has been evaluated *excellent* by *almost all* important experts. But we can also see in figure 3 that this option has so much vague assessment that perhaps we should choose a different machine as the optimal one.

		Alternatives in Υ according to	
$\mathcal{A}_r$	$\mathcal{Q}_s$	original approach	new approach
excellent	**almost all**	**2**	$\emptyset$
excellent	more than half	1,2	$\emptyset$
good	**almost all**	1,2,4	**3,4**
good	more than half	1,2,4	1,3,4
acceptable	almost all	1,2,3,4	1,3,4
acceptable	more than half	1,2,3,4	1,2,3,4
excellent	about half	1,2,4	$\emptyset$
good	about half	1,2,3,4	1,3,4

Table 1 Alternatives belonging to $\Upsilon_{r,s}$

On the contrary, we can see that according to the new approach, *machine3* and *machine4* are pre-selected to be optimal since they have been evaluated *at least good* by *almost all* important experts. The selection of the optimal alternative from these two alternatives can be based, for instance (see section 3), on the group fuzzy evaluations of alternatives that are calculated as the fuzzy weighted averages with the expert's competencies as fuzzy weights. These group fuzzy evaluations can be compared by their centers of gravity. But we can also employ other procedures for comparing these two alternatives.

6 Conclusion

In this paper we have analyzed the 'soft' consensus model proposed by Sukač et al. [6] and suggested to modify the way of computation of the value of truth of the statement that an alternative is evaluated at least in a certain degree with respect to the opinion of at least a certain quantity of important experts. We have shown that this modification is more appropriate than the previous way proposed in [7] as it takes into account also the uncertainty of the fuzzy evaluations of alternatives according to the experts. We have illustrated the problem by the numerical example from economical practice.

Acknowledgements

Supported by the grant No. IGA_PrF_2018_024 Mathematical Models of the Internal Grant Agency of Palacký University Olomouc.

References

[1] Kacprzyk, J. & Fedrizzi, M. (1986). 'Soft'consensus measures for monitoring real consensus reaching processes under fuzzy preferences. *Control and Cybernetics* (**15**) 309–323.

[2] Kacprzyk, J., Fedrizzi, M. & Nurmi, H. (1997). 'Soft'degrees of consensus under fuzzy preferences and fuzzy majorities. In J. Kacprzyk et al. (eds.), *Consensus under Fuzziness*, (pp. 55–83). New York: Springer Science+Business Media.

[3] Klir, G.J. & Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall PTR.

[4] Pavlačka, O. (2011). Modeling uncertain variables of the weighted average operation by fuzzy vectors. *Information Sciences* (**181**) 4969–4992.

[5] Stoklasa, J. et al. (2015). Soft consensus model under linguistically labelled evaluations. In D. Martinčík, J. Ircingová & P. Janeček (Eds.), *Proceedings of the 33rd International Conference on Mathematical Methods in Economics* (pp. 743–748). Plzeň: Západočeská Univerzita.

[6] Sukač, V., Talašová, J. & Stoklasa, J. (2016). A linguistic fuzzy approach to the consensus reaching in multiple criteria group decision-making problems. *Acta Universitatis Palackianae Olomucensis, Facultas Rerum Naturalium, Mathematica* (**55**) (pp. 133–150). Olomouc:

[7] Sukač, V., Talašová, J. & Stoklasa, J. (2016). 'Soft'consensus in decision-making model using partial goals method. In A. Kocourek & M. Vavroušek (Eds.), *Proceedings of the 34th International Conference on Mathematical Methods in Economics MME 2016* (pp. 791–796). Liberec: Technical University of Liberec.

[8] Talašová, J. (2003). *Fuzzy metody vícekriteriálního hodnocení a rozhodování*. Olomouc: Vydavatelství UP, (in Czech).

[9] Zadeh, L., A. (1965). Fuzzy Sets. *Information and control*, **8**(3) 338–353.

[10] Zadeh, L., A. (1975). The concept of a linguistic variable and its application to approximate reasonning-I. *Information Sciences*, **8**(3) 199–249.

Modeling of survivability of enterprises created under the Human Capital Operational Programme - a case study

Anna Szmit[1], Maciej Szmit[2]

Abstract. The researches of the ability to remain alive or continue to exist are conducted in various fields, starting from medicine and ecology, through actuarial science, military science and engineering, up to economy and management. The subject of the research, the results of which will be presented in the article, is survivability of over 500 microenterprises created in Lodz region (Poland) during 2007-2013 years under the Human Capital Operational Programme with the support of the Fundacja Inkubator from Lodz. The article is a continuation and extension of our previous research and covers the next two years of the presented companies' lifecycles and presents results of comparison of fitting of several models (Burr, Pareto, Gompertz–Makeham and Holt) to the empirical data.

Keywords: Survival Analysis, Gompertz-Makeham Model, Burr Distribution, Double Exponential Smoothing, Enterprises' Survival.

JEL Classification: M13
AMS Classification: 62M10

1 Introduction and literature review

According to the Schumpeter's creative destruction concept (see: [26]), the process of destroying enterprises and creating the new ones, is an immanent feature of economic growth, additionally when comparing the survival of enterprises in various economies, it is needed to consider a number of factors, such as the economic condition, current economic policy and even cultural determinants in particular countries.

Small enterprises, including microenterprises, have been considered as the engine of economic growth and for sustainable and equitable development, however duration of their lifetime is usually short, and the dynamics of their creation/destruction is relatively high.

The researches of enterprises ability to remain alive are conducted in various fields; especially there are many publications concerning several aspects of micro-, small and medium enterprises survivability (see e.g. [22], [2], [15]), particularly devoted to critical success factors (see e.g. [1], [14], [22]) and possible relationships between survivability and characteristics of particular enterprise, such as business size (see e.g. [4]), branch of economy (see e.g. [11], [8], [10], [17], [25]) or characteristics of entrepreneur, such as education, experience, sex or age (see e.g. [5], [20], [19], [24]). Enterprises' survivablity modeling bases on methods known from others disciplines such as demography (see e.g. [12], [30]), actuarial sciences (see e.g. [9], [7]) and population ecology (see e.g.[27], [18]). There are a lot of models used for explaining the dynamic of different populations.

The aim of our research is to find the best way to describe micro-enterprises survival phenomenon. The article is a continuation and extension of our previous research covering lifecycles of over 500 microenterprises created in Lodz region (Poland) during 2007-2013. In the article estimation of parameters of several models and results of comparison of their fitting to the empirical data are presented. We analyzed two classic models: Gompertz–Makeham (see: [13], [23]) and special variant of Pareto type II distribution (Lomax distribution – see [21]) and generalization of Pareto distribution i.e. Burr type XII distribution (see: [6], [3]). Additionally, we decided to test Holt smoothing method (see [16]) as sample model containing an autoregressive component, due to previously identified difficulties in adjusting functions differentiable with respect to time variable.

2 The research subject

The data has been collected from several projects supported by Fundacja Inkubator (FI), http://www.inkubator.org.pl/, which is a non-profit organization located in Lodz and its main goal is promotion of entrepreneurship, especially in starting small business and microenterprises. A larger part of such enterprises is created with support of European Union programs, especially under the Human Capital Operational Pro-

[1] Lodz University of Technology, Faculty of Management and Production Engineering, 266 Piotrkowska Street, Lodz, Poland, anna.szmit@p.lodz.pl.
[2] University of Lodz, Faculty of Management, 22/26 Matejki Street, Lodz, Poland, maciej.szmit@uni.lodz.pl.

gramme, because FI is one of beneficiaries of the program. In our research we analyzed microenterprises created in a few projects under the Measure 6.2 (Support and promotion of entrepreneurship and self-employment) and under the Sub-measure 8.1.2 (Support to adaptation and modernization processes in the region) in the years 2007-2013 in the Lodz Region. Within these projects more than 533 person from Lodz region received grants to start their own businesses. Because one of the conditions for the newly created enterprises was to remain alive for the first 12 months (the survivability is forced to be 100%) we modeled the period starting from 12th month of each enterprise's life, additionally we modeled the survivability in the 5.5 years' time horizon so only the enterprises started form 5.5 years ago were selected. The overall number of analyzed enterprises (the population size) was 355. The detailed description of the investigated population is included in our book [28] and article [29]. The investigated time series was aggregated to 1-month resolution. The survival rates of newly created microenterprises for ones existing for at least 3, 4, 5.5 and 6 years are presented on **Figure 1**.

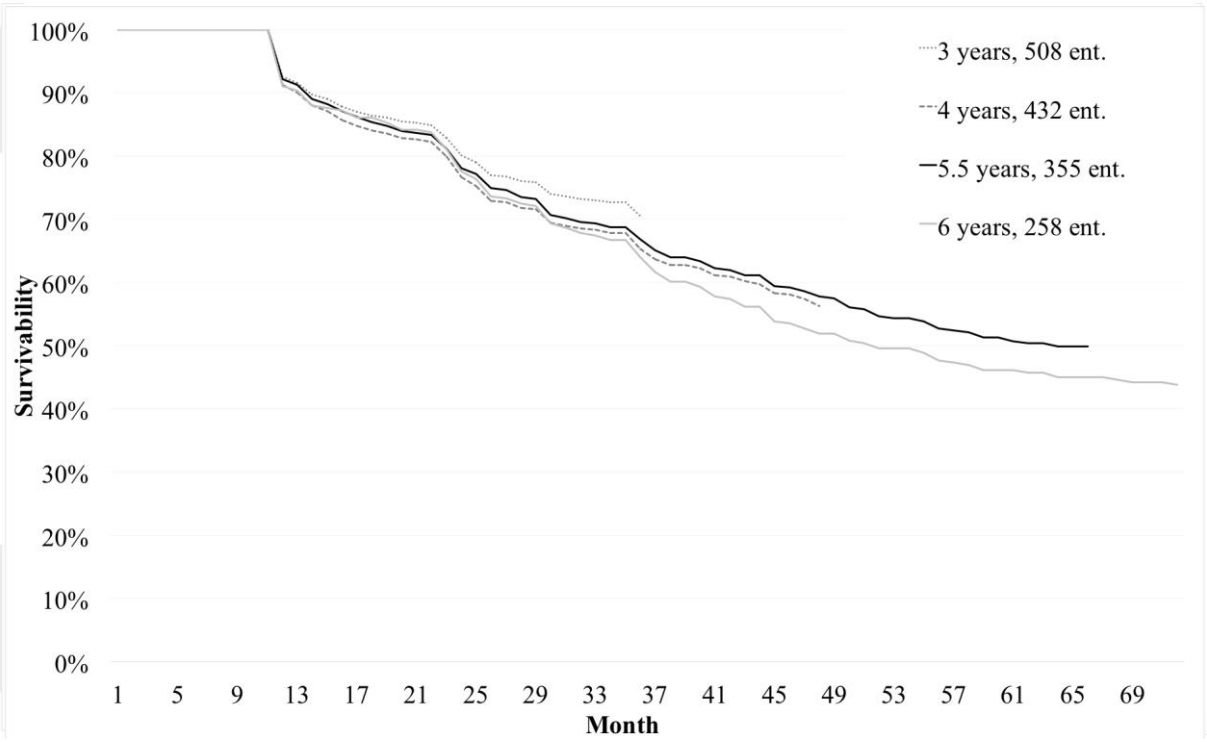


Figure 1. A comparison of survivability of enterprises created under the Human Capital Operational Programme in the Fundacja Inkubator. Source: own research.

3 Methodology

As described above we compared of fitting of several models: Burr, Pareto, Gompertz–Makeham (two variants), Holt to the empirical data.

Gompertz-Makeham model's survival function is given by equation:

$$\hat{S}(t) = \exp\left(\frac{B}{\ln c}(1 - c^t) - At\right)$$

where $\hat{S}(t)$ is estimated value of survivability $S(t)$ in month *t*, *A*, *B* and *c* are parameters, is based on Gompertz–Makeham law of mortality which states that the human death rate is the sum of an age-independent component (Makeham component) and an age-dependent component (the Gompertz function), which increases exponentially with age ($c > 1$). Our research also includes the case where this assumption is skipped, because the enterprises' mortality rate seems not to be increased with age.

Burr Type XII distribution's survival function given by equation:

$$\hat{S}(t) = \left(\frac{b}{b+t^c}\right)^s$$

where: $b>0$, $c>0$, $s>0$ are parameters,

is a composition of Weibull and Gamma distributions.

Pareto distribution's survival function is a special case of the Burr Type XII distribution (c=1), given by the equation:

$$\hat{S}(t) = \left(1 + \frac{t}{b}\right)^{-s}$$

Holt exponential smoothing model (double exponential smoothing) is given by equations:

$$L_t = \alpha\, y_{t-1} + (1-\alpha)(L_{t-1} + T_{t-1}),$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1},$$

$$\hat{S}(t) = \mathrm{L}_t + T_t$$

where:

L_t – the smoothed value for time t,

T_t – estimate of the trend at time t,

α - the data smoothing factor, $0 < \alpha < 1$

and β - the trend smoothing factor, $0 < \beta < 1$.

As a measure of fitting we used Root Mean Square Percentage Error, RMSPE

$$RMSPE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{S(t)-\hat{S}(t)}{S(t)}\right)^2}$$

Because of obligatory selection of initial values in the Holt model its matching for the first observations may not be adequate therefore the period from t = 15 was assumed for determining the average error. For ensuring comparability the same rule was applied for all other models.

As a reference model we considered also the naïve model, where forecasted survivability rate for month n+1 is given by real survivability rate for the previous (n) month.

Parameters of all models were determined using the Method of Least Squares.

4 Results

The results of all models for months 12-66 after enterprise started are presented in the **Table 1**.

Model	**Naïve**	**Burr**	**Pareto**	**Gompertz-Makeham (c > 1)**	**Modified Gompertz-Makeham (c = 0,934)**	**Holt**
RMSPE	1.50%	1.24%	1.31%	4.9%	1.34%	1.17%
first-order residual autocorrelation coefficient	-0.059	0.705	0.753	0.979	0.742	0.130

Table 1. Comparison of fitting of models estimated for months 12-66. Source: own research.

Concerning models' characteristics, the most noticeable is the strong autocorrelation of residuals (except of adaptive models i.e. Holt and the naïve model). This situation is implied by step changes in the time series, what are caused mostly by some tax privileges existing in Polish law system for enterprises during their first two years of live. The models and the original data are presented on the **Figure 2**.

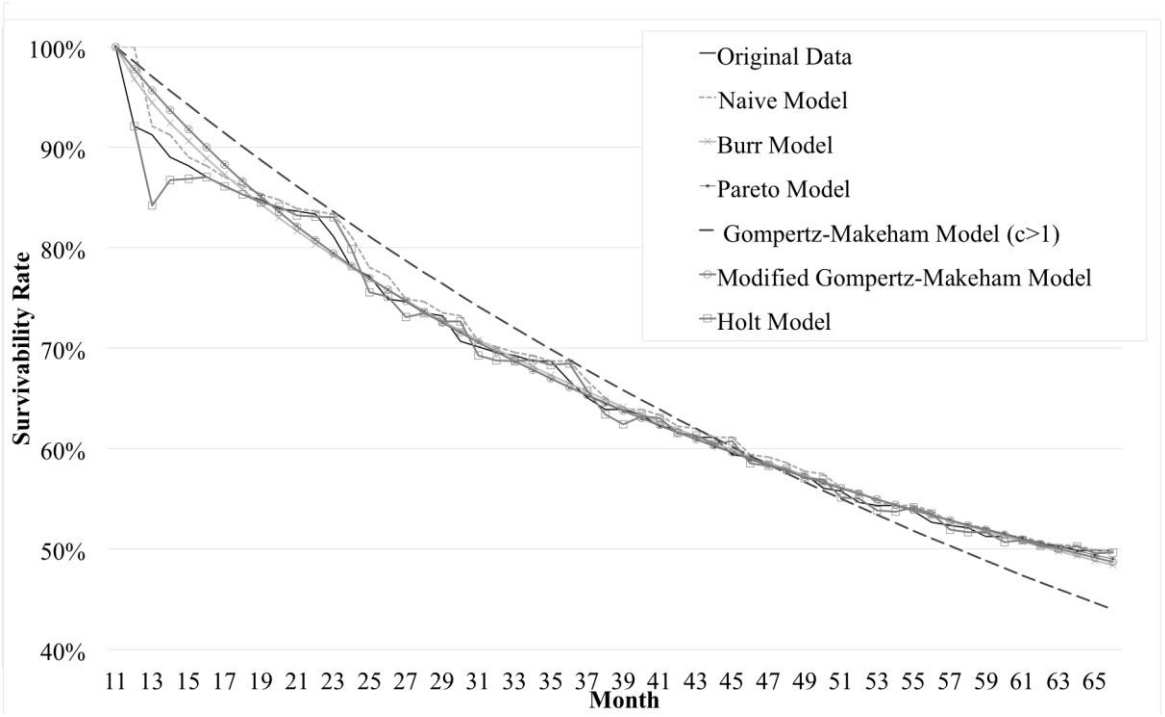


Figure 2. A graphical comparison between original data and model's output. Source: own research.

The Gompertz-Makeham model with classic assumptions ($c > 1$) definitely does not effectively predict the behavior of the timeseries. The Holt model adjusts level and trend factor using exponential smoothing to reflect changes in the examined series. The fitting of curves to empirical data for the other models gives results similar to each other.

We also considered RMSPE for the same models excluding the period before 24th month expecting that models represented by smooth functions should be better suited. Detailed results are presented in the **Table 2**.

Model	Naïve	Burr	Pareto	Gompertz-Makeham (c>1)	Modified Gompertz-Makeham (c=0,972)	Holt
RMSPE	1.55%	0.88%	0,88%	4.64%	0.89%	0.99%
First-order residual autocorrelation coefficient	-0.059	0.390	0.415	0.976	0.445	0.239

Table 2. Comparison of fitting of models estimated for months 24-66. Source: own research.

The results are distinctly better, additionally the autocorrelation of residuals is lower, however it remains significant.

5 Discussion and conclusions

The fitting of all models is relatively high. The best fitting for the full-length time series, with the problem of step changes, was achieved by the model of double exponential smoothing. However, this adaptive model is accurate and relatively robust for such situations, it does not explain the nature of the investigated phenomenon. The other models, without autoregression component, such as Burr, Pareto and Gompertz-Makeham might be useful for better describing and understanding the changes in the investigated phenomenon, however – because of the original laws of mortality were formulated to human population – they may be not fully adequate for modeling enterprises‘ lifecycles. The omission of certain assumptions in the Gompertz-Makeham model in the presented studies gave positive results. It may indicate the occurrence of phenomenon analogous to late-life mortality deceleration, known from other disciplines (see e.g. [12]), however that hypothesis requires more detailed research.

After cutting the second year (months 12-24), containing major step changes from the time series the Holt model gives no longer particularly good fitting. The advantage of Holt method is the ability to adapt to abrupt

changes, after removing them other models fits are better. It can be assumed that this is because these models take into account the character of the modeled phenomenon.

6 References

[1] Aigbavboa C.O., Tshikhudo L.M., Thwala W.D. (2014). Identification of critical success factors for the survival of small, medium and micro enterprise contracting firms in the greater Johannesburg metropolitan area. *Proceedings 8th Construction Industry Development Board*, http://hdl.handle.net/10210/12045

[2] Aremu M. A. Small and Medium Scale Enterprises as A Survival Strategy for Employment Generation in Nigeria. *Journal of Sustainable Development, Vol. 4 (No. 1),* 200-206, https://doi.org/10.5539/jsd.v4n1p200

[3] Balicki A. (2006) *Analiza przeżycia i tablice wymieralności*. Warszawa: Polskie Wydawnictwo Ekonomiczne

[4] Bercovitz, J., Mitchell, W. (2007). When is more better? The impact of business scale and scope on long-term business survival, while controlling for profitability. *Strategic Management Journal, Volume 28*, 61-79, https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.568

[5] Brüderl, J., Preisendörfer P., Ziegler R. (1992). Survival Chances of Newly Founded Business Organizations. *American Sociological Review, Volume 57*, 227-241, https://www.jstor.org/stable/2096207?seq=1#page_scan_tab_contents

[6] Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics Volume 13(2)*, 215–232, doi:10.1214/aoms/1177731607. JSTOR 2235756

[7] Carriere J.F. (1992). Parametric models for life tables. *Transactions of the Society of Actuaries. Vol.44*, https://www.soa.org/library/research/transactions-of-society-of-actuaries/1990-95/1992/january/tsa92v446.pdf

[8] Fertala, N. (2008). The shadow of death: do regional differences matter for firm survival across native and immigrant entrepreneurs? Empirica, *Journal of European Economics, Volume 35 (1)*, 59–80, https://link.springer.com/article/10.1007/s10663-007-9051-2

[9] Forfar D. O. (2006). *Mortality laws, Encyclopedia of Actuarial Science*, John Wiley & Sons, Ltd, https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470012505.tam029 (2006).

[10] Freeman J., Carroll G. R., Hannan M. T. (1983). The Liability of Newness: Age Dependence in Organizational Death Rates, *American Sociological Review, Volume 48 (5)*, 692-710, https://www.jstor.org/stable/pdf/2094928.pdf?seq=1#page_scan_tab_contents

[11] Fritsch M, Brixy U., Falck O. (2004): The Effect of Industry, Region and Time on New Business Survival – A Multi-Dimensional Analysis. *Discussion Papers on Entrepreneurship, Growth and Public Policy*, Jena: Max Planck Institute for Research into Economic Systems Group Entrepreneurship, Growth and Public Policy, ftp://papers.econ.mpg.de/egp/discussionpapers/2004-31.pdf

[12] Gavrilov L.A., Gavrilova N.S., Natalia S. (2011). Mortality Measurement at Advanced Ages: A Study of the Social Security Administration Death Master File, *North American Actuarial Journal, volume 15 (3)*, 432–447, http://longevity-science.org/pdf/Mortality-NAAJ-2011.pdf

[13] 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, 115*, 513-583.

[14] Groenewegen G., de Langen F. (2012). Critical Success Factors of the Survival of Start- Ups with a Radical Innovation, *Journal of Applied Economics and Business Research, Volume 2 (3)*, 155-171, http://www.aebrjournal.org/uploads/6/6/2/2/6622240/4makalecriticalsuccessfac torsofthesurvivalofstart.pdf

[15] Gurgul H., Zając P. (2016). Zastosowanie modelu przyspieszonej porażki do analizy przeżycia małopolskich przedsiębiorstw. *Przegląd Statystyczny, Volume LXIII (4)*, 391-409, http://keii.ue.wroc.pl/przeglad/Rok%202016/Zeszyt%204/2016_63_4_391-409.pdf

[16] Holt, C.C. (1957). *Forecasting trends and seasonals by exponentially weighted averages*, Carnegie Institute of Technology, Pittsburgh ONR memorandum no. 52.

[17] Hyytinena A., Pajarinenb M., Rouvinenb P. (2015). Does innovativeness reduce startup survival rates? *Journal of Business Venturing, Volume 30 (4)*, July 2015, 564–581, http://www.sciencedirect.com/science/article/pii/S0883902614000913

[18] Jørgensen, S.E. (1986), *Fundamentals of Ecological Modelling.* Developments in Environmental Modelling, Amsterdam: Elsevier, vol. 9.

[19] Kalninsa A., Williams M. (2014). When do female-owned businesses out-survive male-owned businesses? A disaggregated approach by industry and geography. *Journal of Business Venturing, Volume 29 (6)*, 822–835, https://www.sciencedirect.com/science/article/pii/S0883902613001092

[20] Kalleberg A. L., Leicht K. T. (1991). Gender and Organizational Performance: Determinants of Small Business Survival and Success, *Academy of Management Journal, Volume 34 (1)*, 136-161, https://www.jstor.org/stable/256305

[21] Lomax, K. S. (1954). Business failures. Another example of the analysis of failure data. *Journal of the American Statistical Association*, Vol. 49, No. 268 (Dec., 1954),847–52. doi:10.1080/01621459.1954.10501239.
[22] Łobos K., Szewczyk M. (2002). Survival analysis: A case study of micro and small enterprises in Dolnośląskie and Opolskie Voivodship (Poland), *Ekonomická revue – Central European Review of Economic Issues, Volume 15*, 207-216, https://www.ekf.vsb.cz/export/sites/ekf/cerei/cs/Papers/VOL15NUM04PAP01.pdf
[23] Makeham, W. (1860). On the Law of Mortality and Construction of Annuity Tables. *The Assurance Magazine and Journal of the Institute of Actuaries, 8(6)*, 301-310
[24] van Praag, C. M. (2003). Business Survival and Success of Young Small Business Owners. *Small Business Economics, Volume 21 (1)*, 1-17, https://link.springer.com/article/10.1023%2FA%3A1024453200297
[25] Shane S. (2012). Small Business Failure Rates by Industry: The Real Numbers, *Small Business Trends, Volume 2012 (9)*, http://smallbiztrends.com/2012/09/failure-rates-by-sector-the-real-numbers.html
[26] Schumpeter J. A. *Capitalism, Socialism and Democracy* (5th ed.), Taylor & Francis e-Library, 2003, http://cnqzu.com/library/Economics/marxian%20economics/Schumpeter,%20Joeseph-Capitalism,%20Socialism%20and%20Democracy.pdf
[27] Siler, W. (1979), A Competing-Risk Model for Animal Mortality. *Ecology, 60*: 750-757, https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1936612
[28] Szmit, A., Lisiak-Felicka, D., Szmit, M. (2017). *Badania przeżywalności firm utworzonych w ramach projektów realizowanych z Programu Operacyjnego Kapitał Ludzki. Studium empiryczne dla Fundacji Inkubator*, Łódź: Wydawnictwo Politechniki Łódzkiej, http://repozytorium.p.lodz.pl/bitstream/handle/11652/1484/ASzmit_DLisiak-Felicka_MSzmit_Badania_przezywalnosci_firm.pdf?sequence=1&isAllowed=y
[29] Szmit M., Szmit A., Dąbkowska M., Lisiak-Felicka D. (2017). An Analysis of Selected Correlations Between Microenterprises' Financing and Its Characteristics. A Case Study. In Wilimowska Z., Borzemski L., Świątek J. (eds) *Information Systems Architecture and Technology: Proceedings of 38th International Conference on Information Systems Architecture and Technology – ISAT 2017*. 141-149. Cham: Springer, https://doi.org/10.1007/978-3-319-67223-6_14
[30] Tabeau E. (2001) A Review of Demographic Forecasting Models for Mortality. In: Tabeau E., van den Berg Jeths A., Heathcote C. (eds) *Forecasting Mortality in Developed Countries European Studies of Population, vol 9*. Springer, Dordrecht, DOI https://doi.org/10.1007/0-306-47562-6_1

Modelling of slaughter pigs' prices in the Czech Republic using stochastic models

Ondřej Šimpach[1], Marie Šimpachová Pechrová[2]

Abstract. Despite that the economic situation in pig meat sector improved in years 2015 and 2016, second half of 2017 was worse again due to decrease of slaughter pig price that tends to be highly volatile. Knowing the future price of slaughter pigs can minimize the risks related to the decision-making at pig farms. Hence, the aim of the paper is to find optimal model that can be used for modelling and predicting of monthly prices of slaughter pigs. Data from 01/1998 to 06/2016 are used to predict slaughter pig price for period 07/2016–06/2017. Two types of stochastic models are used. First, the development of slaughter pig meat producers' prices is modelled by Seasonal Autoregressive Integrated Moving Averages (SARIMA) model. Consequently, the relations with monthly prices of fodder plants (available only for period 02/2006–06/2016) are searched using Vector Autoregressive (VAR) model.

Keywords: SARIMA, slaughter pig price, VAR.

JEL Classification: C61, C63
AMS Classification: 62H12

1 Introduction

The sector of pork meat production has been in crisis in the Czech Republic (CR) recently. The decline of the state of sows is visible since in 1989 due to high competitiveness and cheap imports from abroad. The situation worsened after the entrance of the CR to European Union (EU). Cancelling the import tariffs lead to increased imports of live animals and pig meat. The producer's prices were low and the costs of production high [16]. The most important changes since the entrance to the EU happened in the size of the pig farms. The number of farms reduced, but their size has increased. This enabled them to adjust their production and achieve the returns to scale [10]. Also the price changes of the inputs (feed) cause distortions. For example, during 2014 there has been a high volatility in feed prices resulting in high prices for both cereals and compound feeding stuffs (see [9]). The price of slaughter pigs is volatile, which affects the decision making on the farm. "Extreme price spike and volatility in agricultural commodity prices creates negative effects on macroeconomic instability, posing a threat to food security in many countries," (Bayramoğlu [2]). Like in other sectors, the price is determined based on the agreement between the seller and the buyer. However, as Jaile-Benitez, Ferrer-Comalat and Linares-Mustarós [8] noted, "this agreement rarely remains stable because it is often reached after unwanted pressures, creating situations of dissatisfaction that involve one of the two parties". Farmers' bargaining power is lower. The reasons for asymmetric price transmission in the agro-food chain was examined e.g. by Backus, Falkowski and Fertö [1] using meta-analysis of existing studies.

Modelling and prediction of prices and finding the relations between the developments of various prices had been a subject of a long-time examination by many authors. For example, Rumánková [13] used Box-Jenkins Autoregressive Integrated Moving Averages (ARIMA) modelling method to project the prices of selected agricultural commodities. She found out that mostly the time series are integrated by the order of 1. Saengwong, Jatuporn and Roan [14] also found that the prices of broilers, cattle, duck and hogs are stationary when their first differences were taken into account. Their study also modelled the cointegration of price elasticity, searched for causalities among prices and forecasted future prices of broiler, cattle, duck and hog in Taiwan by using time series analyses (the unit roots, Johansen cointegration, Granger causality and variance decomposition tests). Šimpach [15] observed prices of honey and sugar and found out that "in some years the honey prices and sugar prices have developed in a similar trend. In this case the change in sugar price had slight delay." How the agricultural commodity prices are influenced by oil prices was studied by Fowowe [7] in South Africa. However, there was no evidence of a long-run relationship. Rafiq and Bloch [12] found that there was a non-linear relationship between oil and most of 25 commodity prices they examined using linear and nonlinear ARDL models for long-run relations and asymmetric Granger causality tests for short-run causalities. Fernandez [6] examined the causalities between four US price indices and 31 commodity prices from 1957 to 2011. She found out that "agricultural raw materials

[1] University of Economics Prague, W. Churchill Sq. 4, 130 67 Prague 3, CZ, ondrej.simpach@vse.cz.
[2] Institute of Agricultural Economics and Information, Mánesova 75, 120 00 Prague 2, CZ, simpachova.marie@uzei.cz.

(cotton, hides, rubber, and wool), beverages (coffee), food (maize, rice, and wheat), minerals, ores and metals (copper), and vegetable oilseeds and oils (groundnut oil and soybean oil) display bidirectional linear and non-linear feedback effects vis-à-vis price indices", (Fernandez, [6]).

In order to examine the current situation and to project future development of prices of slaughter pigs in the Czech Republic, univariate and multivariate approaches to the time series analysis are used. Individual time series is scrutinized and forecasted by Box-Jenkinson methodology (SARIMA model), and the type of the relations among multiple time series is examined by VAR model. Both models are also used for predictions. The paper is structured as follows. First, methods and data are described. In the article we are concerned not only with modelling of the individual time series, but the relations between particular prices are also examined and future development is projected. Next section presents results of both forecasts and compares them. Last section concludes.

2 Methodology and Data

There are two groups of methods used for the time series analysis – univariate and multivariate. First, the Box-Jenkins [3] Seasonal Autoregressive Integrated Moving Average (SARIMA) analysis is applied on time series of individual series of price of slaughter pigs. Second, the influence of the price of the feed (short-term relationship) is examined using Vector Autoregressive (VAR) model. Both models are diagnostically tested.

First, the time series is examined by Augmented Dickey-Fuller (ADF) test whether it is stationary or non-stationary as the analysis of the relationship between time series by VAR model can be done only if they are integrated of the same order. There are three types of ADF tests elaborated by Dickey and Fuller [5]: with constant and trend, with constant only, and without constant and trend. We applied the last mentioned as (1)

$$\Delta Y_t = \beta\, Y_{t-1} + \sum_{i=1}^{m} \alpha_i Y_{t-i} + \varepsilon_t \tag{1}$$

where ΔY_t is the first difference of the examined variable, t is time, m is the maximum length of the lagged dependent variable, α, β are parameters, and ε_t is a pure white noise error term.

Diagnostic of the type of ARIMA model is done by Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) that are plotted in order to determine the order p of Autoregressive (AR) process and order q of Moving Average (MA) process. Consequently, the appropriate type of the model is identified. AR model, MA model, Autoregressive Integrated Moving Average (ARIMA) model, and Seasonal ARIMA (SARIMA) model in terms of statistically significant parameters and their ability to explain the correlation structure of the process that generated the time series. In case of agricultural prices some seasonality can be expected (due to natural character of the agricultural production). Therefore, SARIMA(P, D, Q) (p, d, q) model is chosen and estimated as (2)

$$Y_t = \beta + \sum_{i=1}^{P} \mathrm{A}_i Y_{t-is} + \sum_{j=1}^{Q} \mathrm{B}_j \varepsilon_{t-js} + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j} \tag{2}$$

where s marks an order of seasonality. Consequently, fitted models are used to predict the future producers' price of slaughter pig in next 12-month period. Also 95% confidence intervals are elaborated.

Possible correlation between the time series of the price of slaughter pigs and the price of the main feed of the (feed wheat, feed barley, feed oat, and feed maize) are assessed next. It is assumed that they have similar trend and are integrated of the same order. They are tested by ADF test (after and before seasonal adjustment) and consequently by Granger test for spurious regression. General VAR(p) model can be written in the form (3)

$$\mathbf{Y}_t = \boldsymbol{\beta} + \sum_{i=1}^{m} \boldsymbol{\Phi}_i \mathbf{Y}_{t-i} + \boldsymbol{\varepsilon}_t \tag{3}$$

where $\boldsymbol{\beta}$ is l x 1 dimensional vector of constants, $\boldsymbol{\Phi}_i$, $i = 1, 2, \ldots, m$ are l x l dimensional non-random matrices of AR parameters and $\boldsymbol{\varepsilon}_t$ is l-dimensional process of white noise. In our case, price of slaughter pigs is modelled with 5 time series (in natural logarithms). The middle projection and projection with 95% confidence intervals are done in both models for 12-month period.

The presence of autocorrelation is tested using Breusch-Godfrey serial autocorrelation LM test with H_0: There is no serial autocorrelation. If the calculated value of the test exceeds the tabled test criterion from Fisher and χ^2 distribution H_0 is rejected and there is autocorrelation. Heteroscedastity is tested by Autoregressive Conditional Heteroscedasticity (ARCH) test where is H_0: There is no heteroscedasticity present. The test is also using Fisher and χ^2 critical values. Normality is tested by Jarque-Bera test with H_0: The residues are normally distributed and value is compared to critical value of Jarque-Bera distribution. All tests are done at 0.05% level of significance.

Data were taken from database at http://www.agris.cz. Originally, the frequency of slaughter pigs' prices was weekly, but as the prices of feed were available only on the monthly bases, they were transformed on monthly data by calculating the monthly average. There were 222 observations in period from 01/1998 to 06/2016 for slaughter pig price used for SARIMA model. The prices of fodder wheat, fodder barley, fodder oats, and fodder maize were available from 02/2006 to 06/2016 and therefore the VAR model worked with 125 observations only. The projections are compared with real price development from 07/2016 to 07/2017. Calculations were done in EViews 8.

From Figure 1 and Table 1 can be seen that the highest price was around 07/2001 (47.75 CZK/kg). Year 2001 was almost the only one when the prices were higher than 40 CZK/kg. Average price in this year was also the highest from the whole observed period. The prices were fluctuating around 32.33 CZK/kg (arithmetic mean). Over half time, the price was higher than 32.10 CZK/kg, but in 25% of cases lower than 29.32 CZK/kg or higher than 35.07 CZK/kg. On the other hand, the lowest price (23.42 CZK/kg) was noted in 03/1999, but on average the lowest prices were so far noted during the first six months of year 2016. Almost similar situation was in year 2010. The prices usually decrease on the turn of the New Year (only exceptions are 1998/1999, 2002/2003, 2010/2011). "At the beginning of each year, price of pork usually decreases as many households save after the Christmas. Part of consumers is also trying to fulfil New Year's resolutions about loss of weight and healthy lifestyle and the consumption and the prices are decreasing." (Brož, [4])

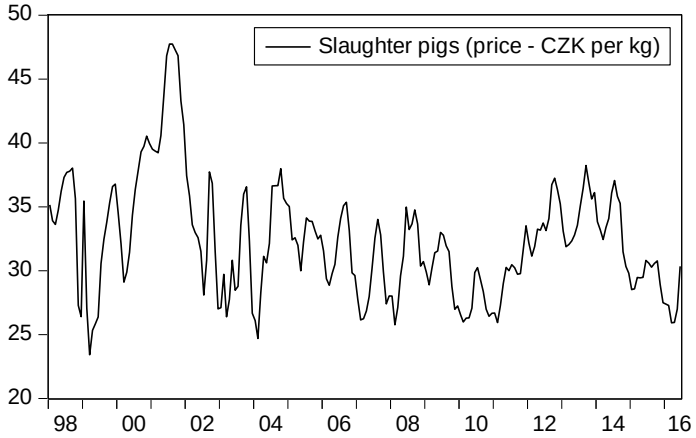


Figure 1 Development of slaughter pigs price [CZK/kg] (01/1998–*06/2017); Source: own elaboration

Year	Price	Year	Price	Year	Price	Year	Price
1998	34.48	**2003**	30.34	**2008**	31.08	**2013**	34.51
1999	30.75	**2004**	32.65	**2009**	30.34	**2014**	33.57
2000	35.43	**2005**	32.87	**2010**	27.51	**2015**	29.58
2001	43.62	**2006**	31.65	**2011**	29.55	**2016**	30.31
2002	33.00	**2007**	29.16	**2012**	34.01	**2017***	33.48

Table 1 Development of slaughter pigs price [CZK/kg] (01/1998–*06/2017); Source: own elaboration

Regarding the prices of fodder wheat, barley, oats and maize, they developed almost equally during the observed period. They were increasing since 2006 until 2008, when due to the financial crisis experienced steep decrease. The highest average prices were in year 2013. In the first half of 2016 they stabilized on higher level than a decade ago. The trend of the development of average prices is almost similar to the development trend of slaughter pigs price, hence, there might be correlation and certain type of short-term relationship.

3 Results and Discussion

First, the prices of slaughter pigs were modelled individually by Box-Jenkinson methodology – SARIMA model and further used for prediction. Second, VAR model was estimated and used for price prediction.

The price of slaughter pigs was modelled individually by Box-Jenkins methodology after the diagnostics by ACF and PACF functions. The results of the model together with verification tests are displayed in Table 2. F-test revealed that the time series is seasonal. After testing, it was found that the most suitable model is SARIMA(1,0,1)(0,0,1)c. The time series is stationary and with autoregressive term AR of the degree of 1. The price depends on the price one month ago. MA is of the first order in non-seasonal version of the model and of the first order in seasonal version. Autocorrelation of residues was also rejected by Durbin-Watson and Breusch-Godfrey tests, but the residues were not normally distributed and constant and finite, hence, the HAC errors were used.

Variable	Coeff. (Std. error) Prob.		
β	32.139 (1.217)***		
AR(1)	0.8060 (0.061)***		
MA(1)	0.3870 (0.074)***		
SMA(12)	0.4530 (0.143)***		
Model diagnostic		**Breusch-Godfrey Serial Correlation LM Test:**	
R^2	0.846	F-statistic	1.267
Adjusted R^2	0.844	Prob. $F^{[2,215]}$	0.284
F-statistic	397.039***	Obs* R^2	2.572
Durbin-Watson statistic	2.049	Prob. $\chi^{2\,[2]}$	0.276

*** statistically significant at $\alpha = 0.01$

Table 2 Price of slaughter pigs: SARIMA(1,0,1)(0,0,1)c model Source: own elaboration

The results of projection are displayed at Figure 2. The maximum price is expected to be 32.16 CZK/kg and in 10/2016 and minimum price 30.02 CZK/kg in 03/2017. Upper bound of the 95% confidence interval follows the trend of the middle prediction. There is also a peak in 10/2016 (39.20 CZK/kg) and minimum in 03/2017 (38.02 CZK/kg). Lower bound shows deep decrease to 30.02 CZK/Kg in 03/2017, but such low price has not been achieved in at any time during the observed period and hence it is not probable.

VAR model examined short-time relationship between the development of producers' prices of slaughter pigs and their possible determinants: the price of fodder wheat, barley, oat and maize. Fisher test for seasonality found that barley is non-seasonal, wheat, oat and maize are seasonal and were seasonally adjusted. Individual ADF models with constant revealed that each time series is non-stationary and stationary of the first order. Residues of Granger spurious regression test are stationary at 0.05 level of significance. Hence, the regression among variables is not spurious. All prices of fodder plants Granger cause the price of slaughter pigs. Therefore, all are included in VAR model and we obtain VAR model of the 6th grade (lag = 6). Results of the VAR model are presented in Table 3. There is no serial autocorrelation of the residues and no heteroskedasticity, but due to length of time series, the residues are not normally distributed.

The price of slaughter pigs depends on itself at first and third lag. Then it is dependent on the price of maize with two and three lags, price of oat at first, second lag, price of wheat at first, fifth and even sixth lag, price of barley at first and third lag. All fodder prices depend on the price of fodder wheat at first lag. Wheat as it can be stored for long time can also influence the prices up to half a year later (fourth, fifth, sixth lag). Price of oat is important only up to third lag, and maize up to forth lag. The least important is price of barley that depends only on its price one month ago and influences only the price of slaughter pigs. Impulse was statistically significant in case of all fodder plants with exception of wheat. This dummy variable took value of 1 at 06/2013 when there was a significant change in price. The price of fodder oat was almost twice higher as month before (4518 CZK/t) and after (4427 CZK/t). Price of maize was lower (4292 CZK/t) in 06/2016 than in 05 (5373CZK/t) or 07 (5210 CZK/t). Price of barley decreased from 5008 CZK/t in 05 to 4766 CZK/t in 06 and continued to decrease. Similar situation was with maize that decreased from 5505 CZK/t in 05 to 5306 CZK/t in 06/2013.

Predicted values o slaughter pig price are given at Figure 2. The middle variant suggests slight increase to 30.50 CZK/kg in 08/2016, but the price will decrease again on 29.50 CZK/kg in 10/2016. Then it should slightly increase on 29.94 CZK/kg at 12/2016 and decrease again on 29.54 CZK/kg one month later. Then there is no significant change, only slight increase on 29.93 CZK/kg at the end of the projected period 06/2017. Lower bound suggests pessimistic development that the price will decrease from 28.62 CZK/kg down to 24.58 CZK/kg in 06/2017. In 10, 11/2016 the price shall be the same 26.11 CZK/kg and then increase on 26.29 CZK/kg. However, further decrease is predicted. Upper bound is an optimistic variant starting from 31.58 CZK/kg in 07/2016 and increasing up to 35.29 CZK/kg. There shall be only little stagnation of the trend between 08-10/2016. Otherwise the price should continually rice.

Middle variants predicted by SARIMA and VAR model suggest slight increase and then slight decrease of the price of slaughter pigs to the minimum. The prices finally rise at the end of the period. In case of SARIMA model, the price might raise to 32.16 CZK/kg in 10/2016 and 32.00 CZK/kg in 06/2017. VAR model predicts the highest price 30.50 CZK/kg in 08/2016. It is a general rule that if in the VAR model, the variables are Granger caused by each other, the confidence intervals are smaller. The confidence intervals in VAR models are almost two times narrower than in the case of SARIMA model. While the difference between prices projected by SARIMA is between 7.08 and 16.02 CZK/kg, VAR model predicts the difference only between 2.96 and 10.71 CZK/kg. Hence, the theoretical expectations were confirmed.

Prices of	pigs	maize	oat	wheat	barley
Variable	Coeff. (SE) Prob.	Coeff. (SE) Prob.	Coeff. (SE) Prob.	Coeff. (SE) Prob.	Coeff. (SE) Prob.
pigs (-1)	1.000 (0.108)***	31.055 (38.941)	-12.052 (30.589)	33.208 (29.789)	-15.811 (29.681)
pigs (-3)	-0.324 (0.149)**	66.890 (53.579)	-24.511 (42.087)	31.396 (40.987)	19.488 (40.839)
maize (-1)	-0.001 (0.000)	0.321 (0.113)***	-0.031 (0.089)	-0.002 (0.087)	-0.133 (0.086)*
maize (-2)	-0.001 (0.000)**	0.262 (0.111)***	0.085 (0.087)	0.527 (0.085)***	0.056 (0.084)
maize (-3)	0.001 (0.000)**	0.006 (0.136)	-0.037 (0.106)	-0.493 (0.104)***	-0.090 (0.103)
maize (-4)	-0.000 (0.000)	-0.066 (0.143)	0.253 (0.112)**	0.160 (0.109)*	-0.084 (0.109)
oat (-1)	0.001 (0.000)***	0.369 (0.122)***	0.530 (0.096)***	0.079 (0.093)	0.185 (0.093)**
oat (-2)	0.001 (0.000)***	0.149 (0.113)*	0.248 (0.089)***	0.292 (0.086)***	0.026 (0.086)
oat (-3)	-0.000 (0.000)	-0.323 (0.124)***	-0.182 (0.097)**	-0.336 (0.095)***	-0.091 (0.094)
wheat (-1)	-0.001 (0.000)***	0.484 (0.142)***	0.386 (0.112)***	0.998 (0.109)***	0.264 (0.108)***
wheat (-4)	0.000 (0.001)	0.007 (0.191)	-0.040 (0.150)	-0.143 (0.146)	0.519 (0.145)***
wheat (-5)	-0.001 (0.001)*	-0.010 (0.241)	-0.610 (0.190)***	0.051 (0.185)	-0.639 (0.184)***
wheat (-6)	0.002 (0.001)***	0.154 (0.230)	0.485 (0.165)***	0.097 (0.161)	0.218 (0.160)*
barley (-1)	-0.001 (0.000)*	0.009 (0.153)	0.106 (0.120)	0.130 (0.117)	0.501 (0.116)***
barley (-3)	0.001 (0.001)***	-0.041 (0.164)	0.128 (0.129)	0.069 (0.125)	-0.047 (0.125)
IMP (dummy)	-3.806 (1.534)***	-1068.603 (551.874)**	1505.897 (433.504)***	176.288 (422.170)	-648.559 (420.642)*
Model diagnostic					
R^2	0.934	0.925	0.928	0.958	0.957
Adj. R^2	0.912	0.899	0.903	0.944	0.942
F-statistic	41.568	35.995	37.723	67.084	64.657

Table 3 Price of slaughter pigs: VAR model; Source: own elaboration

Lower bound of SARIMA prediction suggests pessimistic development that the price will decrease even on 22.02 CZK/kg in 03/2017, but it has never been so low before. VAR model minimum is 24.58 that is more realistic. Upper bound of SARIMA model supposes the increase from 34.77 CZK/kg in 07/2016 up to 40.01 CZK/kg eleven months later while VAR model projects the highest price of 35.29 CZK/kg. Both models are realistic, as the maximum price achieved was 47.75 CZK/kg, despite that it was already in year 2001. The difference between projected and real values was the lowest in case of upper bound of VAR model confidence interval. The second-best projection measured by the variance was middle SARIMA projection.

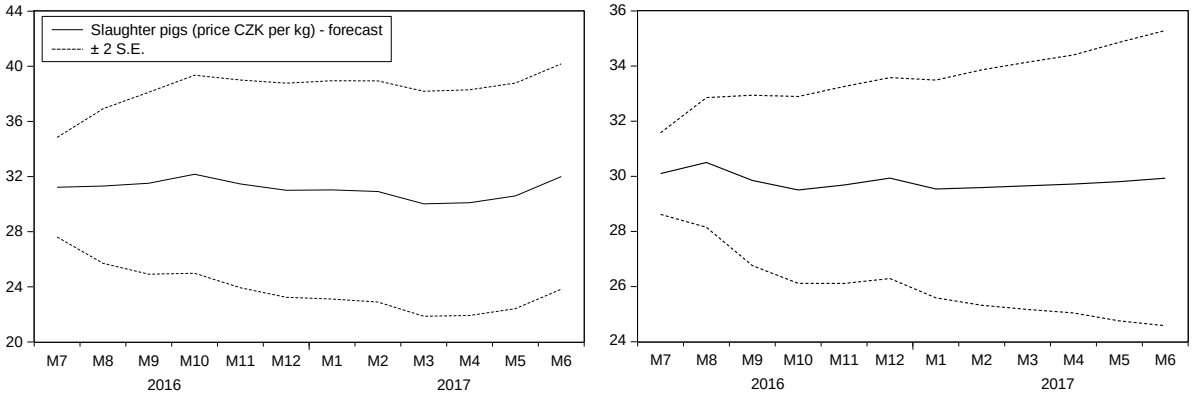


Figure 2 Predictions of producers' prices of slaughter pigs based on SARIMA model (left) and VAR model (right); Source: own elaboration

Models of Box-Jenkins methodology, ARCH or VAR models are suggested by many researchers to be used for projection of prices of agricultural commodities (see e.g. [11]). The research of Jaile-Benitez, Ferrer-Comalat and Linares-Mustarós [8] can provide comparable results to ours. The authors searched for the determinants (explanatory variables) of price of the pork and tried to quantify their weights (degree of influence) using fuzzy logic. They found that the variable influencing the price and their weights are cyclical and suffer variations in the course of events. We suggest using VAR model as it provided better results even when shorter time series was available. Besides, using more related time series enables to test the causality between variables (see e.g. study of Saengwong, Jatuporn and Roan [14]).

4 Conclusion

Analysis of the agricultural products' price volatility and trend forecasting are necessary to formulate and implement business strategies of agricultural holdings and for policy-making. Therefore, the aim of the paper was to find the optimal model for modelling and predictions of monthly prices of slaughter pigs. Producers' price was modelled by Autoregressive Integrated Moving Averages (ARIMA) models (time series from 01/1998 to 06/2016)

and by Vector Autoregressive (VAR) model in relation with prices of fodder plants (data from 02/2006 to 06/2016). Predictions are done for 12 months (until 06/2017).

The lower bound of SARIMA model of slaughter pigs' prices is too pessimistic, while according to VAR, the results are more probable. Middle and upper bound projections are again lower in case of VAR model, but feasible in both cases. SARIMA predicts the middle variant in interval 30.02 CZK/kg to 32.16 CZK/kg and VAR from 29.50 CZK/kg to 30.50 CZK/kg. Despite that the time series in the case of VAR models is eight years shorter than in case of SARIMA, the prediction in case of slaughter pigs is realistic and usable and we suggest using VAR model, where the upper bound values were the closest to the reality. Also, the variance of middle projection was not high. However, the second lowest difference between reality and projection was found in case of middle projection done by ARIMA model.

Acknowledgement

The research was supported by the Czech Science Foundation project no. P402/12/G097 DYME – Dynamic Models in Economics and from Internal Research Project no. 1110 of Institute of Agricultural Economics and Information.

References

[1] Bakucs Z., Falkowski, J., Ferto, I. (2014). Does Market Structure Influence Price Transmission in the Agro-food Sector? A Meta-Analysis Perspective. *Journal of Agriculture Economics, 65(1),* 1–25.

[2] Bayramoğlu, A. T. (2014). The Impact of Agricultural Commodity Price Increases on Agricultural Employment in Turkey. *Procedia - Social and Behavioral Sciences, 143*, 1058–1063.

[3] Box, G. E. P., Jenkins, G. (1970). *Time series analysis: Forecasting and Control*. San Francisco: Holden-Day, 575 p.

[4] Brož, J. (2014). Ceny vepřového se propadly. Vánoční hostina může být levnější. *In: IDNES.CZ.* Available at http://ekonomika.idnes.cz/vykupni-ceny-veproveho-jsou-na-nejnizsi-urovni-od-roku-2010-p3h-/test.aspx?c=A141204_200742_test_ozr (accessed Nov 1, 2016)

[5] Dickey, D. A., Fuller, W. A. (1979). Distribution of Estimators for Autoregressive Time Series with a Unit Root. *Journal of American Statistical Association, 74(366),* 427–431.

[6] Fernandez, V. (2014). Linear and non-linear causality between price indices and commodity prices. *Resources Policy, 41*, 40–51.

[7] Fowowe, B. (2016). Do oil prices drive agricultural commodity prices? Evidence from South Africa. *Energy, 104(1),* 149–157.

[8] Jaile-Benitez, J. M., Ferrer-Comalat, J. C., Linares-Mustarós, S. (2015). Determining the Influence Variables in the Pork Price, Based on Expert Systems. In Gil-Aluja J., Terceño-Gómez A., Ferrer-Comalat J. C., Merigó-Lindahl J. M., Linares-Mustarós S. (eds.), *Scientific Methods for the Treatment of Uncertainty in Social Sciences, 377*, 81–92.

[9] Marquer, P., Rabade, T., Forti R. (2014). Pig farming in the European Union: considerable variations from one Member State to another. (Pig farming sector - statistical portrait 2014). *Statistics in focus, 15*. Available at http://ec.europa.eu/eurostat/statistics-explained/index.php/Pig_farming_sector_-_statistical_portrait_2014 (accessed Nov 1, 2016).

[10] Pechrová M., Medonos T. (2016). Total factor productivity change in the Czech poultry sector. In *The 10th International Days of Statistics and Economics (MSED 2016).* Slaný: Melandrium, 1806–1815.

[11] Pechrová, M., Šimpach, O. (2017). Modelling the development of eggs prices in the CR. In *Proceedings of 15th International Scientific Conference on Hradec Economic Days 2017*, University of Hradec Kralove, vol. 7(1), 682–688.

[12] Rafiq, S., Bloch, H. (2016). Explaining commodity prices through asymmetric oil shocks: Evidence from nonlinear models. *Resources Policy, 50*, 34–48.

[13] Rumánková, L. (2016). 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: CULS Prague, 319–326.

[14] Saengwong, S., Jatuporn, C., Roan, S. W. (2012). An Analysis of Taiwanese Livestock Prices: Empirical Time Series Approaches. *Journal of Animal and Veterinary Advances 11(23),* 4340–4346.

[15] Šimpach O. (2012). Statistical view of the current situation of Beekeeping in the Czech Republic. In *The 6th International Days of Statistics and Economics (MSED 2012).* Slaný: Melandrium, 1054–1062.

[16] Štolcová, J., Homolka, J. (2012). Production Costs in the Fattening Period of Pigs and Assessment of the Influence of Selected Factors on the Amount of These Costs. *Agris on-line Papers in Economics and Informatics, 4(4 – Special),* 107–113.

The order of contacts to be called after an emergency situation - fuzzy multicriteria decision-making approach

Michal Škoda[1], Helena Brožová[2]

Abstract. This article deals with a quantitative support of decision-making in emergency management in order to select and organize emergency contacts according to their importance in relation to the possible consequences of an emergency situation. The decision support model is primarily designed to meet the needs of gas distribution dispatch centres. However, it can be applied to other sectors as well. With regard to the limited information that the employee of dispatch centre has available immediately after the emergency, the entire system is designed to work with vague input data. The used model is designed as a multicriteria decision-making model with fuzzy evaluation.

The initial step of the process starts with the evaluation of the severity of all possible threats using fuzzy scale. For each threat a list of possible contacts with their importance is created. When an emergency situation occurs, the severity of the threats resulting from the current situation is evaluated on the fuzzy scale. Then the synthesis of the default and current evaluation of the threats importance take a place. After that the importance of individual contacts is calculated and lexicographically arranged. The summary of these steps is the overall functional model for the decision support.

Keywords: Fuzzy number, linguistic scale, contacts, emergency situation, dispatch centre.

JEL Classification: C44, D70
AMS Classification: 90B50, 62C86

1 Introduction

According to Steiner, Nussbaumer and Albert [4], decision-making is a key aspect of managing emergency situations and a successful response to an emergency situation depends highly on whether decisions are taken in an effective and timely manner. Emergencies represent a decision-making environment characterized mainly by high complexity due to the large number and interdependence of variables, dynamics, uncertainty and non-transparency of the whole situation, excessive information or lack of information, time pressure, risk, plurality of goals and involvement of multiple entities. Therefore, it is useful to work with fuzzy mathematics and multi-criteria decision making methods.

Sometimes a really critical situation may occur during which even an extremely capable and experienced employee make a mistake. For instance, who would you call first during a cyber attack on a gas high-pressure transfer station? National Office for Cybernetics and Information Security? Network manager? Fire Department? Of course, dispatchers and other key employees are trained for such situations, but an emergency is an emergency. One never knows what can happen, how he/she will react, whether he/she will panick, whether he/she correctly evaluate all criteria, etc. The model that is described below aims to help the dispatcher and other employees in such situations, to streamline the situation and guide the employee to its evaluation. Thanks to this evaluation the model creates a structured list of emergency contacts that should be most appropriate for the situation. The whole model will work with vague terms to make it as user-friendly as possible. The vague description of an emergency event is usually all that dispatchers have at the first moments of an emergency.

According to Ye, Dai and Qiu [6], the use of the fuzzy set theory for emergency response has been proven to be useful as well as the use of decision support systems with the flexible organization of their functions and reuse of the functions. These topics are now one of the central topics in theoretical research and applications in the decision support system for an emergency situation. Yoon, Velasquez, Partridge and Nof [7] also state that the decision support systems will be a key tool for improving the efficiency of the emergency management. This article wants to contribute to this trend.

[1] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamycka 129, Prague, Czech Republic, mskoda@pef.czu.cz.
[2] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamycka 129, Prague, Czech Republic, brozova@pef.czu.cz.

In the environment of gas distribution companies, the first one who finds out about an emergency situation is usually a dispatcher, who should evaluate the situation and respond adequately to it. It also means that all the relevant persons and authorities, such as for example members of the emergency commission, the fire brigade, the Energy Regulatory Office (ERO), etc., should be informed as soon as possible. Model which is described in this article should be able to help a dispatcher to make a quick and effective decision regarding emergency contacts to call and in what order.

The aim of this article is to design a simple model based on fuzzy tools and multicriterial decision making method, with output of a set of emergency contacts, ranked by importance in relation to the specific emergency situation.

The structure of the article is as follows: the second chapter describes the used tools of fuzzy mathematics and the method of multicriteria decision-making. The third chapter describes the proposed model (a basis of proposed decision support system), and the fourth chapter provides an example of a specific use of the system. In conclusion, the proposed model and results are evaluated.

2 Method for decision-making in emergency situations

The proposed model is based on multicriteria decision-making method, especially on the lexicographic method and the evaluation using fuzzy linguistic scales.

2.1 Fuzzy Linguistic Scale

Within the evaluation of the individual elements of the model the different types of linguistic fuzzy scales will be used. There is a huge number of different linguistic scales. The main division of linguistic scales is into even scales and uneven scales. Furthermore, the scales can be divided into two-part scale, three-part scale, etc. [8].

For the purposes of classifying some specific emergency situations a six-part uneven scale will be used. The reason for choosing this particular scale is to avoid the center value. This corresponds with a large gap between the 3rd and 4th point (Figure 1). The intention is to divide the values into "worse than center" and "better than center". The details about the six-point scale are given in the Table 1 and Figure 1.

Linguistic term	Shorcuts	Fuzzy number
Very High	VH	(0,9, 1, 1, 1)
High	H	(0,7, 0,8, 0,9, 1)
Moderately High	MH	(0,4, 0,6, 0,7, 0,8)
Moderately Low	ML	(0,2, 0,3, 0,4, 0,6)
Low	L	(0, 0,1, 0,2, 0,3)
Very Low	VL	(0, 0, 0, 0,1)

Table 1 Six-point scale

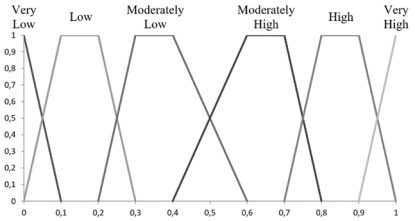


Figure 1 Six-point scale

The scale to be used to categorize emergency contacts, more precisely to identify the importance of individual contacts depending on the threat involved, is a five-point uneven scale. The scale was chosen based on the assumption that people in their evaluation tend to move in their choices to the center of the scale and try to avoid extreme values. The membership function is slightly modified for this scale, taking into account that if the evaluator chooses an extreme value, he is presumed to be more certain by his choice than the evaluator who chooses the mean value. This means that on this scale the greatest uncertainty have the center values and with the approach to extreme values, the uncertainty decreases and at the extreme value is the uncertenty minimal [8]. The graphical representation of the scale is shown in Figure 2. This five-point scale will also be used to determine the importance of each threat type. The five-point scale will include the following degrees of importance of the contacts and their corresponding fuzzy number (Table 2, Figure 2).

Linguistic term	Shorcuts	Fuzzy number
Very Important	VI	(0,8, 0,9, 1, 1)
Important	I	(0,6, 0,7, 0,8, 0,9)
Moderately Important	MI	(0,3, 0,4, 0,6, 0,7)
Slightly Important	SI	(0,1, 0,2, 0,3, 0,4)
Not Important	NI	(0, 0, 0,1, 0,2)

Table 2 Five-point scale

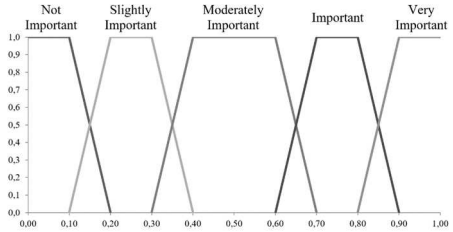


Figure 2 Five-point scale

The proposed fuzzy scales will be represented by trapezoidal fuzzy numbers $A(x) = (x_1, x_2, x_3, x_4)$, where $x_1 \leq x_2 \leq x_3 \leq x_4$. For the membership function $A(x)$ applies [2]:

$$A(x) = \begin{cases} 0 & if \quad a < x_1 \\ \dfrac{a - x_1}{x_2 - x_1} & if \quad a \in [x_1, x_2] \\ 1 & if \quad a \in [x_2, x_3] \\ \dfrac{x_4 - a}{x_4 - x_3} & if \quad a \in [x_3, x_4] \\ 0 & if \quad a > x_4 \end{cases} \tag{1}$$

For the model purposes we need to add and multiply fuzzy numbers and rank fuzzy numbers. If we have two fuzzy numbers A and B the addition of fuzzy numbers A and B is defined by the formula [5]:

$$(A +)(z) = sup\{min\{A(x), B(y)\} | x + y = z, x, y \in U\} \tag{2}$$

Multiplication of fuzzy numbers A and B is defined by the formula [5]:

$$(A \times B)(z) = sup\{min\{A(x), B(y)\} | x \times y = z, x, y \in U\} \tag{3}$$

The ranking of fuzzy numbers within the proposed model is based on the ranking of their centers of gravity. Suppose $A(x)$ is a trapezoidal fuzzy number that is determined by four points (x_1, x_2, x_3, x_4). Then for the center of gravity $t_{A(x)}$ of fuzzy number $A(x)$, the following applies:

$$t_{A(x)} = \frac{1}{3} \cdot \frac{x_4^2 + x_3^2 - x_2^2 - x_1^2 + x_4 x_3 - x_2 x_1}{x_4 + x_3 - x_2 - x_1} \tag{4}$$

If A and B are fuzzy numbers and if $t_A \leq t_B$ holds, the fuzzy number A is smaller than fuzzy number B due to their center of gravity, we say that $A \leq B$ [5].

2.2 Lexicographical ranking method

The lexicographic method is generally based on the principle that the greatest influence on the choice of the most important variant has the most important criterion. Only when there are several variants that are equally ordered according to the most important criterion, then the second most important criterion is considered. If we cannot organize variants by the second most important criterion, then comes the third most important criterion etc. The algorithm stops when a complete arrangement is found or when all the criteria are used [1].

2.3 Fuzzy multicriteria model for emergency contact ranking

Determination of general significance of individual types of threat

The first and essential step in this model is the identification of basic types of threats for the possible emergency scenarios. For the purposes of the dispatchers of the gas distributors dispatch centers the selected types of threats may be for example life threat, health threats, property threat, environmental threats etc.

In the second step it is necessary to determine the general severity of each type of threat by using the five-point fuzzy scale described in Table 2. The importance of each type of threat may change frequently. For example, the threat of gas supply has a different importance in summer, when the gas consumption is generally lower, than in winter, when the gas consumption is significantly higher thanks to the heating plants and boiler rooms.

Determining the importance of individual contacts within individual threat types

The next step consist of identification of the relevancy of the emergency contacts, which the company usually has in document such as Crisis Preparedness Plan or the Emergency Plan. The importancy of these selected emergency contacts must always be evaluated within a specific threat by using the five-point scale described in Table 2.

Classification of a specific emergency situation

In case that an emergency situation take a place, the specific severity of each type of threat has to be evaluated. This evaluation is made by using the six-point fuzzy linguistic scale described in Table 1.

An alternative evaluation of the importance of individual types of threat can be done directly in one step within the classification of a specific emergency situation, but it would be less user-friendly, because the decision-maker has to to take into account several different aspects at once.

Ranking of contacts

Once the individual contacts are assigned to the fuzzy categories, the significance of the individual types of threat are determined and a specific emergency situation is classified, we have everything necessary to create a ranked list of emergency contacts.

Step 1: The first step is to determine the overall preference of the threat synthezing the general and the specific severity of the individual threats. This is done by multiplying (Formula 3) the general severity of a type of threat and the specific severity of a type of threat given by the specific emergency situation.

The overall preference of the individual threats can also serve as an indicator of the criticality of the given situation. The higher maximum value of the center of gravity (Formula 4) of the overall threats severity is, the more serious the emergency situation is and vice versa.

Step 2: The second step is calculation of the importancy of emergency contacts within the particular threat multiplying the calculated overall threat severity with individual contact preferences. The result will be preferences expressed as fuzzy numbers assigned to each contact according to the individual threat types. To order these emergency contacts, it is necessary to convert their fuzzy evaluation to real numbers. This is done by calculating their center of gravity (Formula 4).

Step 3: The final arrangement of emergency contacts based on their importance in terms of individual threats is based on the lexicographic method, which is based on the ranking of the importance of the individual contacts across the individual threats. All the centres of gravity of fuzzy evaluation of each emergency contact are first sorted from the largest to the smallest for each contact. Now the emergency contacts are ranked based on the ranked center of gravity values using lexicographic method. If the result is only partial ranking, the contacts with the same preferences are considered to be equivalently important and therefore their order can be determined randomly [3].

Model structure

The Table 3 provides the overview of model construction and application. The table is divided into three sections. Into the first one (light grey) the initial general evaluation or data are filled in. The second part (medium gray) contains the specific evaluation of the actual emergency situation. The third part of the model (dark grey) consist of model results.

		Threats							
		T1	T2	T3	...				
Initial evaluation	General severity								
	Emergency contacts								
	C1								
	C2								
	...								
Specific evaluation	**Emergency situation**								
	Specific severity								
Model calculation	Overall severity								
	Emergency contacts					Lexicographic ranking			
	C1								
	C2								
	...								

Table 3 Model schema

3 Case study

The designed model for ranking a list of emergency contacts immediately after the beginning of an emergency situation was applied to the environment of one of the three major gas distributors in the Czech Republic. For the purposes of this article the fictive scenario below was given:

A construction company disrupt a high-pressure gas pipeline. The gas started to leak and burn intensively in an inhabited area.

Based on the scenarios available to the dispatching center of the selected distribution company, seven basic types of threats have been identified. The degree of general severity of identified types of threats has been determined by employees involved in crisis management through fuzzy terms (Table 4).

Type of Threat	Life Threat	Health Threat	Threat to Gas Supplies	Threat to Property	Cyber Threat	Enviromental Threat	Threat to Reputation
	LT	HT	TGS	TP	CT	ET	TR
General severity	Very Important	Very Important	Important	Important	Moderately Important	Moderately Important	Slightli Important
Fuzzy Scale	(0,8, 0,9, 1, 1)	(0,8, 0,9, 1, 1)	(0,6, 0,7, 0,8, 0,9)	(0,6, 0,7, 0,8, 0,9)	(0,3, 0,4, 0,6, 0,7)	(0,3, 0,4, 0,6, 0,7)	(0,1, 0,2, 0,3, 0,4)

Table 4 Identified threats and their general severity

Base on study of the company's internal documents, a total of 25 key emergency contacts were identified, which, depending on their nature, were assigned to individual types of threats and were evaluated in cooperation with the distribution company. For instance, in Table 5 the first ten selected emergency contacts are listed with their importance with regard to appropriate threat type.

Emergency contacts	LT	HT	TGS	TP	CT	ET	TR
Call centre	-	-	I	-	I	-	I
Czech Environmental Inspectorate	-	-	-	-	-	SI	-
Czech Radio	-	-	SI	-	-	-	MI
Member of Emergency Commission I.	SI	SI	I	I	I	SI	MI
Member of Emergency Commission II.	SI	SI	I	I	I	SI	MI
Member of Emergency Commission III.	SI	SI	I	I	I	SI	MI
Dispatch Center of innogy	-	-	SI	-	SI	-	-
ERO	-	-	MI	-	-	-	SI
Fire Brigade	VI	VI	VI	VI	-	VI	-
Office of the Board	NI	NI	NI	NI	NI	NI	MI

Table 5 Identification, assigment and evaluation of importancy of the selected contacts

Step 1: The given emergency situation was evaluated by the employee of the company as follows in Table 6.

Type of Threat	Life Threat	Health Threat	Threat to Gas Supplies	Threat to Property	Cyber Threat	Enviromental Threat	Threat to Reputation
General severity	Moderately High	Very High	High	High	Very Low	Low	High
Fuzzy Scale	(0,4, 0,6, 0,7, 0,8)	(0,9, 1, 1, 1)	(0,7, 0,8, 0,9, 1)	(0,7, 0,8, 0,9, 1)	(0, 0, 0, 0,1)	(0, 0,1, 0,2, 0,3)	(0,7, 0,8, 0,9, 1)

Table 6 Specific severity of the emergency situation

Step 2: The overall threat preference is calculated for the individual threats based on the specific situation. The calculation of this preference is based on fuzzy terms, respectively based on their fuzzy numbers. As stated in Chapter 2.3, this will be done by multiplying the general fuzzy evaluation of the type of threat and specific fuzzy evaluation of the threat for the specific situation. The resulting severity are listed in the Table 7:

Type of Threat	Life Threat	Health Threat	Threat to Gas Supplies	Threat to Property	Cyber Threat	Enviromental Threat	Threat to Reputation
Fuzzy Scale	(0,32, 0,54, 0,70, 0,80)	(0,72, 0,90, 1,00, 1,00)	(0,42, 0,56, 0,72, 0,90)	(0,42, 0,56, 0,72, 0,90)	(0,00, 0,00, 0,00, 0,07)	(0,00, 0,04, 0,12, 0,21)	(0,07, 0,16, 0,27, 0,40)

Table 7 The overall severity of the emergency situation

Step 3: Subsequently, the overall preference for each threat is multiplied by the importance of contacts that belong to the appropriate threat type. The resulting fuzzy numbers pertaining to the individual contacts are defuzzyfyed through the center of gravity. The results of this step are shown on the 5 most important and 5 less importatnt contacts in Table 8.

	According to the threats severity							According to the centers of gravity values						
Contacts	LT	HT	TGS	TP	CT	ET	TR	MAX1	MAX2	MAX3	MAX4	MAX5	MAX6	MAX7
Fire Brigade	0,556	0,839	0,615	0,615	-	0,093	-	0,839	0,615	0,615	0,556	0,093	-	-
Medical Service	0,556	0,839	-	-	-	-	-	0,839	0,556	-	-	-	-	-
Emergency Commission	0,461	0,687	0,615	0,615	0,023	0,093	0,219	0,687	0,615	0,615	0,461	0,219	0,093	0,023
Gas Emergency Service	0,461	0,687	0,615	0,615	0,016	0,081	-	0,687	0,615	0,615	0,461	0,081	0,016	-
Police	0,556	0,687	-	0,615	-	-	-	0,687	0,615	0,556	-	-	-	-
⋮														
Call centre	-	-	0,511	-	0,021	-	0,186	0,511	0,186	0,021	-	-	-	-
OTE	-	-	0,356	-	-	-	0,135	0,356	0,135	-	-	-	-	-
ERO	-	-	0,356	-	-	-	0,072	0,356	0,072	-	-	-	-	-
Ministry of Industry	-	-	0,356	-	-	-	0,072	0,356	0,072	-	-	-	-	-
ÚAMK	0,169	0,238		0,067	-	-	0,072	0,238	0,169	0,072	0,067	-	-	-

Table 8 Ranked contacts with the centers of gravity values

The final ranked list of the first 10 contacts that arose based on the lexicographic ranking of the centres of gravity is given in Table 9.

1.	Fire Brigade	6.	Vice-Chairman of Emergency Commission
2.	Medical Service	7.	Member of Emergency Commission I.
3.	Chairman of Emergency Commission	8.	Member of Emergency Commission II.
4.	Gas Emergency Service	9.	Member of Emergency Commission III.
5.	Police	10.	Spokesman of the Company

Table 9 The final order of the first ten contacts

4 Conclusion

The model that generates the ranked list of contacts fulfilled the goal and the expectations of the employees involved in emergency management in the company.
The illustrative example and its results confirms the usability of the Operational Research tools in very specific decision-making areas as emergency management. It is shown that fuzzy evaluation using appropriately chosen fuzzy linguistic scales and multi-criteria approach to choosing the most appropriate decision are very useful and provide the desired results.
According to the interviewed senior employees, the model may be suitable especially for new and less experienced employees. The necessary time to open the model and fill in the required data was the main weakness for them. Based on this remark, we can say that with increasing numbers of contacts, which may be used during an emergency and with variability in their arrangement, the benefit arising from using this model is growing. On the other hand, despite this remark, we can say that the showed example confirms the possibility of using this model in practice as a basis of potential decision-support system in the early stages of the emergency situation.

Acknowledgements

This research is supported by the grant No. 20181012 “Systémová podpora manažerského rozhodování v krizových situacích” of the Internal Grant Agency of the University of Life Sciences Prague.

References

[1] Hwang, C.L. & Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications: a state-of-the-art survey.* New York: Springer-Verlag. ISBN 0387105581.
[2] Klir, G. J. & YUAN, Bo. (1995). *Fuzzy sets and fuzzy logic: theory and applications.* Upper Saddle River, N.J.: Prentice Hall PTR. ISBN 978-0131011717.
[3] Roudný, R. & Víšek, O. (2009). *Základy manažerského rozhodování.* Pardubice: Univerzita Pardubice, 184 s. ISBN 978-80-7395-164-1.
[4] Steiner, Ch., Nussbaumer, A. & Albert, D. (2015). A psychological framework modelling decision making in Emergencies. *Proceedings of 9th European Conference on IS Management IS Management and Evaluation*, 206-214. ISSN 2048-8912.
[5] Talašová, J. (2003). *Fuzzy metody vícekriteriálního hodnocení a rozhodování.* Olomouc: Univerzita Palackého. ISBN 80-244-0614-4.
[6] Ye, X., Dai, Z. & Qiu, J. A metadata-based flexible function model of decision support system for emergency management. *Proceedings of 4th International Conference on Wireless Communications, Networking and Mobile Computing: WiCOM '08*; Dalian, China, 12 - 17 October 2008. Piscataway, NJ: IEEE Operations Center, 2008, 1-4. ISBN 9781424421084.
[7] Yoon, S. W., Velasquez, J. D., Patridge, B. K. & Nof, S. Y. (2008). Transportation security decision support system for emergency response: a training propotype, *Decision Support Systems*, 139-148.
[8] Zemková, B. (2012). *Využití fuzzy množin v oblasti řízení lidských zdrojů.* Olomouc. Disertační práce (Ph.D.). Univerzita Palackého v Olomouci. Přírodovědecká fakulta.

Two Algorithms for Risk-averse Reformulation of Multi-stage Stochastic Programming Problems

Martin Šmíd[1], Václav Kozmík[2]

Abstract. Many real-life applications lead to risk-averse multi-stage stochastic problems; therefore effective solution of these problems is of great importance. Many tools can be used to their solution (GAMS, Coin-OR, APML or, for smaller problems, Excel); it is, however, mostly up to researcher to reformulate the problem into its deterministic equivalent. Moreover, such solutions are usually one-time, not easy to modify for different applications.
We overcome these problems by providing a front-end software package, written in C++, which enables to enter problem definitions in a way close to their mathematical definition. Creating of a deterministic equivalent (and its solution) is up to the computer.
In particular, our code is able to solve linear multi-stage with Multi-period Mean-CVaR or Nested Mean-CVaR criteria. In the present paper, we describe the algorithms, transforming these problems into their deterministic equivalents.

Keywords: Multi-stage stochastic programming, deterministic equivalent, multi-period CVaR, nested CVaR, optimization algorithm.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Generally, adding risk measures into decision problems destroys their favourable properties such as convexity and linearity. The CVaR risk measure, however, is an exception because it may be reformulated as a convex minimization problem [2], which can be linearised if the underlying distribution is discrete with finite number of atoms. It follows that the favourable properties of CVaR are inherited by the Mean-CVaR risk criterion (see e.g. [4]). Consequently, Mean-CVaR decision problems with convex or linear payoff and constraint functions may be reformulated as convex ones, linear ones, respectively. In the present paper, we do the same in the dynamic case. In particular, we show how to preserve convexity/linearity of multi-stage stochastic programming problems with the Multi-period Mean-CVaR criterion or the Nested Mean-CVaR criterion (see [3] for more about these and other dynamic risk measures).

2 Multistage Stochastic Programming Problem

Let $T \geq 0$. We define T+1-stage Stochastic Programming Problem as

$$\min_{x_k \in \mathcal{X}_k 0 \leq k \leq T} \quad \rho\left(f_0(\overline{\xi}_0, \overline{x}_0), \dots, f_T(\overline{\xi}_T, \overline{x}_T)\right), \tag{1}$$

$$\mathcal{X}_k = \mathcal{X}_k(\overline{\xi}_k, \tilde{x}_{k-1}) = \left\{ x_k \in \mathbb{R}^{d_k}, x_k \in \mathcal{F}_k : x_k \in R_k(\overline{\xi}_k), g_k(\overline{\xi}_k, \overline{x}_k) \underset{\geq_k}{\leqq} 0 \text{ a.s.} \right\}.$$

Here, $\xi = (\xi_0, \dots, \xi_T)$ is a random process taking values in $\mathbb{R}^{n_0} \times \dots \times \mathbb{R}^{n_T}$ with ξ_0 deterministic, $\mathcal{F}_0, \dots, \mathcal{F}_T$ is its induced filtration. Further, for each $0 \leq k \leq T$, d_k is a deterministic constant. For each $0 < k \leq T$, we define $\overline{x}_k = (\tilde{x}_{k-1}, x_k)$, where $\tilde{x}_{k-1}$ is a sub-vector of $\overline{x}_{k-1}$, and we put $\overline{x}_0 = x_0$; symbol $\overline{\xi}_k$ is defined analogously. For each $0 \leq k \leq T$ and each feasible $\overline{x}_k$, $f_k(\bullet, \overline{x}_k)$ is a measurable function such that $\mathbb{E}|f_k(\overline{\xi}_k, \overline{x}_k)| < \infty$, and g_k is a r_k-vector of functions, each of which is non-constant in x_k and measurable in $\overline{\xi}_k$; here, $r_k \in \{0, 1, \dots\}$ is $\overline{\xi}_k$-measurable. Further, for each $0 \leq k \leq T$,, $\underset{\geqq_k}{\leqq}$ is a $\overline{\xi}_k$-measurable r_k-vector of symbols from $\{=, \leq, \geq\}$ and R_k is a Cartesian product of d_k closed $\overline{\xi}_k$-measurable intervals. Finally, ρ is a real (risk) functional on the space of integrable random variables which can be the expected sum

$$\rho = \rho^{\mathbb{E}}, \qquad \rho(Z_0, \dots, Z_T) = \sum_{k=0}^{T} \mathbb{E} Z_k,$$

[1]Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, Praha 8, Czech Republic, smid@utia.cas.cz
[2]Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, Praha 8, Czech Republic

the Multi-period Mean-CVaR

$$\rho = \rho^m_{\lambda,\alpha} \qquad \rho^m_{\lambda,\alpha}(Z_0,\ldots,Z_T) = Z_0 + \sum_{k=1}^{T} \rho_{\lambda,\alpha}(Z_k|\mathcal{F}_{k-1}), \qquad \rho_{\lambda,\alpha}(Z|\mathcal{F}) = \left[(1-\lambda)\mathbb{E}(Z|\mathcal{F}) + \lambda \mathrm{CVaR}(Z|\mathcal{F})\right],$$

where $0 \leq \lambda, \alpha \leq 1$, or the Nested Mean-CVaR

$$\rho = \rho^n_{\lambda,\alpha} \qquad \rho^n_{\lambda,\alpha}(Z_0,\ldots,Z_T) = \rho_{\lambda,\alpha}(\rho_{\lambda,\alpha}(\ldots \rho_{\lambda,\alpha}(\sum_{k=0}^{T} Z_k|\mathcal{F}_{T-1})\ldots\mathcal{F}_1)|\mathcal{F}_0),$$

where $0 \leq \lambda, \alpha \leq 1$.

3 Reformulation of the Risk-Averse Problems

If $\rho = \rho^{\mathbb{E}}$ and $f_0,\ldots,f_T$ are convex or linear, then so is ρ. Specially, if ξ is discrete with a finite number of atoms and all the f's and g's are linear, then the deterministic equivalent of (1) is a linear programming problem.

If $\rho = \rho^m$ then, by Proposition 1 (i) (see Appendix), problem (1) may be reformulated as

$$\min_{x_{0,\ldots,T} \in \mathcal{X}_{0,\ldots,T}, u_{0,\ldots,T-1} \in \mathcal{F}_{0,\ldots,T-1}} \quad f_0(\overline{\xi}_0, \overline{x}_0) \quad + \mathbb{E}\left(\sum_{k=1}^{T} h_k(\overline{\xi}_k, \overline{x}_k, u_{k-1})\right) \tag{2}$$

where $h_k(\overline{\xi}_k, \overline{x}_k, u_{k-1}) = u_{k-1} + \left[f_k(\overline{\xi}_k, \overline{x}_k) - u_{k-1}\right]_{\mu,\nu}$; here, $[x]_{a,b} = \begin{cases} xa & x \leq 0 \\ xb & x \geq 0 \end{cases}$, $\mu = 1-\lambda$ and $\nu = 1 - \lambda + \frac{\lambda}{\alpha}$. If $f_1,\ldots,f_T$ are convex, so are $h_t,\ldots,h_T$ (note that $[x]_{\mu,\nu}$ is convex). Moreover, if the distribution of ξ is discrete finite, we apply Proposition 1 (ii) to reformulate (2) as

$$\min_{x_{0,\ldots,T} \in \mathcal{X}_{0,\ldots,T}, u_{0,\ldots,T-1} \in \mathcal{F}_{0,\ldots,T-1}, \theta_{1,\ldots,T} \in \Theta_{1,\ldots,T}} f_0(\overline{\xi}_0, \overline{x}_0) + \mathbb{E}\left(\sum_{k=1}^{T} \theta + \sum_{k=0}^{T-1} u_k\right),$$

$$\Theta_k = \Theta_k(\overline{\xi}_k, \overline{x}_k, u_{k-1}) = \{\theta_k \in \mathbb{R}, \theta_k \in \mathcal{F}_k : \theta_k \geq [f_k(\overline{\xi}_k, \overline{x}_k) - u_{k-1}]_{\mu,\nu}\}, \quad 1 \leq k \leq T. \tag{3}$$

Specially, if f_k and g_k are linear, $k = 0,\ldots,T$, then the deterministic equivalent of (3) is a linear programming problem.

Finally, if $\rho = \rho^n_{\lambda,\alpha}$ and ξ is discrete with a finite number of atoms such that, without loss of generality, the support of ξ_k is $\{1,2,\ldots,m_k\}$ for any k, and that $\mathbb{P}[\xi_{k+1} \in \bullet | \xi_0,\ldots,\xi_k] = \mathbb{P}[\xi_{k+1} \in \bullet | \tilde{\xi}_k]$ (the latter implying $\rho_{\lambda,\alpha}(h(\overline{\xi}_k)|\mathcal{F}_{k-1}) = \rho_{\lambda,\alpha}(h(\overline{\xi}_k)|\tilde{\xi}_{k-1})$ for any h by the definition of Mean-CVaR), then, by the translational invariance of CVaR and by Proposition 1 (iii),

$$\begin{aligned}
\min_{x_\bullet \in \mathcal{X}_\bullet} \rho(f_0,\ldots,f_k) &= \min_{x_\bullet \in \mathcal{X}_\bullet} \left[f_0 + \rho_{\lambda,\alpha}(f_1 + \rho_{\lambda,\alpha}(f_2 + \cdots + \rho_{\lambda,\alpha}(f_T|\tilde{\xi}_{T-1})\ldots|\tilde{\xi}_1))\right] \\
&= \min_{x_0 \in \mathcal{X}_0} \{f_0 + \rho_{\lambda,\alpha}(\min_{x_1 \in \mathcal{X}_1} \{f_1 + \rho_{\lambda,\alpha}(\min_{x_2 \in \mathcal{X}_2} \{f_2 + \cdots + \rho_{\lambda,\alpha}(\min_{x_T \in \mathcal{X}_T} \{f_T\}|\tilde{\xi}_{T-1})\ldots|\tilde{\xi}_2)\}|\tilde{\xi}_1)\})\} \\
&= \min_{x_0 \in \mathcal{X}_0} q_0(\xi_0, x_0)
\end{aligned}$$

where, for any $0 \leq k \leq T-1$,

$$q_k(\overline{\xi}_k, \overline{x}_k) = f_k(\overline{\xi}_k, \overline{x}_k) + \rho_{\lambda,\alpha}\left(Q_k(\overline{\xi}_{k+1}, \tilde{x}_k)|\tilde{\xi}_k\right), \qquad Q_k(\overline{\xi}_{k+1}, \tilde{x}_k) = \min_{x_{k+1} \in \mathcal{X}_k(\overline{\xi}_{k+1}, \tilde{x}_k)} q_{k+1}(\overline{\xi}_{k+1}, \overline{x}_{k+1}),$$

and $q_T(\overline{\xi}_T, \overline{x}_T) = f_T(\overline{\xi}_T, \overline{x}_T)$. Moreover, if, for some $1 \leq k \leq T-1$,

$$Q_k(\overline{\xi}_{k+1}, \tilde{x}_k) = \min_{y \in \mathcal{Y}_k^{\xi_{k+1}}(\tilde{\xi}_k, \tilde{x}_k)} \zeta_k^{\xi_{k+1}}(y; \tilde{\xi}_k, \tilde{x}_k) \tag{4}$$

for some functions $\zeta^1, \zeta^2, \ldots, \zeta^{m_{k+1}}$ and parametric sets $\mathcal{Y}^1,\ldots,\mathcal{Y}^{m_{k+1}}$, we get, by Proposition 1 (iii) and (iv), that

$$q_k(\overline{\xi}_k, \overline{x}_k) = f_k(\overline{\xi}_k, \overline{x}_k) + \min_{u \in \mathbb{R}, y^i \in \mathcal{Y}_k^i(\tilde{\xi}_k, \tilde{x}_k), \theta^i \geq [\zeta^i(y^i; \tilde{\xi}_k, \tilde{x}_k) - u]_{\mu,\nu}, i=1,\ldots,m_{k+1}} \left\{u + \sum_{i=1}^{m_{k+1}} \theta^i \pi^i_{k+1}(\tilde{\xi}_k)\right\} \tag{5}$$

where $\pi^i_{k+1}(\tilde{\xi}_k) = \mathbb{P}[\xi_{k+1} = i|\tilde{\xi}_k]$, $1 \le i \le m_{k+1}$, and, consequently,

$$Q_{k-1}(\overline{\xi}_k, \tilde{x}_k) = \min_{(x_k,u,y^1,\theta^1,\dots,y^{m_{k+1}},\theta^{m_{k+1}})\in\mathcal{Y}^{\xi_k}_{k-1}(\tilde{\xi}_{k-1},\tilde{x}_{k-1})} \left\{ f_k(\overline{\xi}_k, \overline{x}_k) + u + \sum_{i=1}^{m_{k+1}} \theta^i \pi^i_{k+1}(\tilde{\xi}_k) \right\}$$

where

$$\mathcal{Z}^{\xi_k}_{k-1}(\tilde{\xi}_{k-1}, \tilde{x}_{k-1}) = \mathcal{X}_k(\overline{\xi}_k, \tilde{x}_{k-1}) \times \mathbb{R} \times \mathcal{Y}^1(\tilde{\xi}_k, \tilde{x}_k) \times \{\theta^1 \ge [\zeta^1(y^1; \tilde{\xi}_k, \tilde{x}_k) - u]_{\mu,\nu}\} \times \\ \cdots \times \{\theta^{m_{k+1}} \ge [\zeta^{m_{k+1}}(y^{m_{k+1}}; \tilde{\xi}_k, \tilde{x}_k) - u]_{\mu,\nu}\}$$

which fulfills (4) with $k-1$ in place of k. As (4) holds for $k = T-1$, we get, by induction, that (5) holds for any $0 \le k \le T-1$. Consequently,

$$\min_{x_\bullet \in \mathcal{X}_\bullet} \rho(f_0, \dots, f_k) = \min_{x_0\in\mathcal{X}_0, u\in\mathbb{R}, y^i\in\mathcal{Y}^i_1(\tilde{\xi}_0,\tilde{x}_0), \theta^1 \ge [\zeta^i_1(y^i;\tilde{\xi}_0,\tilde{x}_0)-u]_{\mu,\nu}, i=1,\dots,m_1} \left\{ f_0(\xi_0, x_0) + u + \sum_{i=1}^{m_1} \theta_i \pi_i \right\} \quad (6)$$

Specially, if $f_0, \dots, f_T$are linear in x and $g_0, \dots, g_T$ affine in x then, by induction, (6) is a linear programming problem.

4 Conclusion

In this paper, two algorithms linearizing risk averse multistage stochastic programming problems were described. These algorithms are implemented in the MS++, which is a C++ software package developed by the authors. Even though the algorithms use standard techniques, which are routinely applied to solve particular problems (see e.g. [1]), they have been neither rigorously described nor generally implemented yet to the best knowledge of the authors. The package is freely available on `https://github.com/cyberklezmer/mspp`.

Acknowledgements

This work was supported by grant No. GA 16-01298S of the Czech Science Foundation. The support is kindly acknowledged.

References

[1] Václav Kozmík and David P Morton. Evaluating policies in risk-averse multi-stage stochastic programming. *Mathematical Programming*, 2014.

[2] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

[3] Alexander Shapiro and Alois Pichler. Time and dynamic consistency of risk averse stochastic programs. *Published electronically in: Optimization Online*, 2016.

[4] Martin Šmíd, František Zapletal, and Jana Hančlová. Which carbon derivatives are applicable in practice? a case study of a european steel company. *Kybernetika*, 53(6):1071–1085, 2017.

Appendix

Proposition 1. *(i) For any integrable random variable Z,*

$$\rho_{\lambda,\alpha}(Z|\mathcal{F}) = \min_{u\in\mathbb{R}} \left[u + \mathbb{E}(\,[Z-u]_{\mu,\nu}\,|\mathcal{F}) \right]$$

where $[x]_{a,b} = \begin{cases} xa & x \le 0 \\ xb & x \ge 0 \end{cases}$, $\mu = 1-\lambda$, $\nu = 1-\lambda+\frac{\lambda}{\alpha}$.

(ii) If $Z|\mathcal{F} \sim (z_i, \pi_i)_{i=1}^n$, then

$$\rho_{\lambda,\alpha}(Z|\mathcal{F}) = \min_{u\in\mathbb{R}, \theta_i \ge \mu(z_i-u), \theta \ge \nu(z_i-u), i=1,\dots,n} \left[u + \sum_{i=1}^n \theta_i \pi_i \right].$$

(iii) If, in addition, $\mathcal{F}$ is finite and $z_i = \min_{x_i\in\mathcal{X}_i} \zeta_i(x_i)$, $i = 1, \dots, n$, for some sets $\mathcal{X}_1, \mathcal{X}_2 \dots$ and functions $\zeta_1, \zeta_2, \dots,$ all possibly dependent on some $\mathcal{F}$-measurable parameter, then

$$\rho_{\lambda,\alpha}(Z|\mathcal{F}) = \min_{x_i\in\mathcal{X}_i, i=1,\dots,n} \{\rho_{\lambda,\alpha}(\zeta_i(x_i))\}$$

(iv) Moreover,

$$\rho_{\lambda,\alpha}(Z|\mathcal{F}) = \min_{u\in\mathbb{R},x_i\in\mathcal{X}_i,i=1,\dots,n}\left\{u+\sum_{i=1}^{n}\left[\zeta_i(x_i)-u\right]_{\mu,\nu}\pi_i\right\}$$
$$= \min_{u\in\mathbb{R},x_i\in\mathcal{X}_i,\theta_i\geq\mu(\zeta_i(x_i)-u),\theta_i\geq\nu(\zeta_i(x_i)-u),i=1,\dots,n}\left[u+\sum_{i=1}^{n}\theta_i\pi_i\right].$$

Proof. Ad (i) and (ii). Clearly, $cx+d\left[x\right]_+ = \left[x\right]_{c.c+d}$ for any $c,d\geq 0$. Using that and [2] and we get

$$\begin{aligned}
\varrho(Z) &= (1-\lambda)\mathbb{E}(Z)+\lambda\mathrm{CVaR}_\alpha\left(Z\right)\\
&= (1-\lambda)\mathbb{E}Z+\lambda\min_u\left(u+\frac{1}{\alpha}\mathbb{E}\left[Z-u\right]_+\right)\\
&= \min_u\left[(1-\lambda)\mathbb{E}Z+\lambda u+\frac{\lambda}{\alpha}\mathbb{E}\left[Z-u\right]_+\right]\\
&= \min_u\left[(1-\lambda)\mathbb{E}Z-(1-\lambda)u+u+\frac{\lambda}{\alpha}\mathbb{E}\left[Z-u\right]_+\right]\\
&= \min_u\left[(1-\lambda)\mathbb{E}(Z-u)+u+\frac{\lambda}{\alpha}\mathbb{E}\left[Z-u\right]_+\right]\\
&= \min_u\left[u+\mathbb{E}\left((1-\lambda)(Z-u)+\frac{\lambda}{\alpha}\left[Z-u\right]_+\right)\right]\\
&= \min_u\left[u+\mathbb{E}\left[Z-u\right]_{\mu,\nu}\right] = \min_{u,\theta_i=[Z_i-u]_{\mu,\nu}}\left[u+\sum_{\mathrm{i}}\pi_i\theta_i\right]\\
&= \min_{u,\theta_i\geq\mu(Z_i-u),\theta_i\geq\nu(Z_i-u_i)}\left[u+\sum_{\mathrm{i}}\pi_i\theta_i\right] = \min_{u,\theta\in\sigma(Z),\theta\geq\mu(Z-u),\theta\geq\nu(Z-u)}\left[u+\mathbb{E}(\theta)\right]
\end{aligned}$$

Ad. (iii) and (iv). Denote $x=(x_1,\dots,x_n)$, $\mathcal{X}=\mathcal{X}_1\times\cdots\times\mathcal{X}_n$ and $\hat{x}=(\arg\min_{x_1\in\mathcal{X}_1}\zeta_1(x),\dots,\arg\min_{x_n\in\mathcal{X}_n}\zeta_n(x))$. For any $x\in\mathcal{X}$, introduce a random variable $Y(x)\sim(\zeta_i(x_i),\pi_i)$. We have $Z\leq Y(x)$ for any $x\in\mathcal{X}$ so, by monotonicity of Mean-CVaR, $\rho(Z)\leq\rho(Y(x))$ so $\rho(Z)\leq\inf_x\rho(Y(x))$. As $Y(\hat{x})=Z$, the infimum is attained, so

$$\rho(Z)=\min_{x\in\mathcal{X}}(\rho(Y(x)).$$

The first equality in (iv) now follows by (i), the second one may be got analogously to the proof of (ii). □

Claim frequency models in vehicle insurance based on GLM

Adéla Špačková[1]

Abstract. Within non-life insurance pricing determinated by insurance premium precedes and essential part represented models of claim frequency and claim severity. Theese models are usually modelling by generalized linear models. This paper is focused on estimation of claim frequency and extends the work ŠPAČKOVÁ, Adéla. Estimation of claim frequency by generalized linear models, 2017, s. 821-830. ISBN 978-80-248-4138-0. Regression analysis allows the identification of the risk factors and the prediction of the expected frequency of claims given the characteristics of policyholders. It depends on many individual rating factors (e.g. based on individual characteristics of vehicle and driver). The aim of this paper is to find out ideally suited model for estimation claim frequency based on these risk factors. All empirical models are estimated on the real-world sample data of czech insurance company collected during the years 2005-2010. Parameters of model are estimated by maximum likelihood method at standard level of significant 0,05. Verification of the model parameters is perfomed by a Wald test. Comparison models with different predictor variables is established by analysis of deviance residuals, Akaike information criterion (AIC) and Bayesian information criterion (BIC). Based on these comparison the ideally suited model is choosen. All calculations are computed in statistical software STATA 14.00.

Keywords: generalized linear models, vehicle insurance, claim frequency, individual rating factors.

JEL Classification: C13, G22
AMS Classification: 97M30

1 Introduction

In today's sophisticated world all activities are exposed to the risk. Risk can be understood as a negative consequence of the randomness, which can influence individuals and the whole world, where the consequences are huge (mostly financial) losses. With the increasing financial literacy of the whole society, they should feel the occasion of protect through negotiation premium.

The randomness of insurance event and their consequence (i.e., the claim frequency and severity) are a source of uncertainty of the future insurance benefits, which are going to pay out to the clients. This uncertainty can be expressed by an appropriate distribution of probability.

2 Model description

The aim of this paper is to find out ideally suited model for estimation claim frequency based on individual rating factors. Claim frequency models can occur on n number of insurance contract during the certain time interval. Claim frequency is included on interval of non-negative cell numbers $p_k = P(Y = k), k = 0,1,...$

where Y is the number of evets (losses, claims) detected in a fixed time period in a homogenous groups of insureds (collective model).

Based on this predication the claim frequency models should be modelling by disctrete distributions on non-negative integer values (counting distributions). As suitable distribution could be negative-binomial distribution or poisson distribution. The main disadvantage of poisson distribution is overdispersion, because of this allegation for the purposes in this part of science the negative-binomial distribution is selected.

2.1 Negative-binomial distribution

Negative-binomial distribution is a discrete probability distribution of the number of successes and can be derived from Bernoulli trial. Bernoulli trial means probability experiment whose outcome variable is random, where two possible options exist (success or failure). The negative binomial distribution is the number of Bernoulli trials in

[1] VŠB-Technical University, department of Finance, Sokolská třída 33, 702 00 Ostrava, adela.spackova@vsb.cz.

excess of *r*th success. In the other word, we consider the number of failures before the occurrence of the *r*th success, where *Y* is random variable. The probability function of *Y* is following:

$$p_k(Y=k)=\binom{k+r-1}{k}\cdot(1-\pi)^k\pi^r, k=0,1,2... \tag{1}$$

where $x+r$ are trials, where first trial $x+r-1$ included k failures.

Mean is defined:

$$\mu=E(Y)=\frac{r(1-\pi)}{\pi}, \tag{2}$$

variance function is denoted:

$$\sigma^2=Var(Y)=\mu(1+\frac{1}{r}\mu) \tag{3}$$

3 Generalized linear models

Generalized linear models (GLM) generalizes classic linear model in two assumptions. A probability of distribution function is a member of exponential family includes the normal, binomial, poisson, gamma and others. A linear predictor is transform by link function *g,* such that:

$$g(\mu)=x'\beta \tag{4}$$

where *g* is link function and μ is mean, more in (GRAY, Roger J. a Susan M. PITTS (2012) a HARDIN, James W. a Joseph HILBE (2012).

3.1 Exponential family

Basic of GLM modelling is probability of distribution, which is includes exponential family. Exponential family of these distributions provides a general framework for selecting a possible parameterization of the distribution. The special form is selected for mathematical convenience, based on some useful algebraic tools. Generally, exponencial family can be denoted in GLM case:

$$f(Y)=c(Y,\varphi)\exp\left\{\frac{Y\theta-(\alpha\theta)}{\varphi}\right\} \tag{5}$$

Where *f(Y)* is function of response variable includes exponencial family, $c(Y,\varphi)$ and $(\alpha\theta)$ are known parameters, which define probability of selected distribution dependent variable *Y*, θ is unknown canonical link, ϕ means dispersion parameter which is unknown too, see in HARDIN, James W. a Joseph HILBE (2012) a LONG, J. Scott a Jeremy FREESE (2014).

3.2 Link function

Link function transforms vector of dependent variable and provides the relationship between linear predictor and mean of distribution function. There are many commonly used link functions, and the choice depends of assumptions selected probability. The log-link function is choosen in this paper, because prediction of claim frequency should be positive variable. Linear log-link function on basic level is following:

$$\ln(\mu)=g(\mu) \text{ and then, } \mu_i=e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}...+\beta_nx_{in}}. \tag{6}$$

3.3 Maximum likelihood estimation

The maximum likelihood method is based on likelihood function typical for estimation parameters in GLM. This method attempts to find out the each independent variable that maximize the likelihood function, given the observations, more in LONG, J. Scott a Jeremy FREESE (2014) and GRAY, Roger J. a Susan M. PITTS (2012). Maximizing the log-likelihood is:

$$l(\beta,\phi)=\sum_{i=1}^{n}\ln f(Y_i;\beta,\phi)=\sum_{i=1}^{n}\left\{\ln c(Y_i,\phi)+\frac{Y_i\theta_i-\alpha(\theta_i)}{\phi}\right\}, \tag{7}$$

Maximum of $l(\beta,\varphi)$ is differentiated with respect to β_j:

$$\frac{\partial l}{\partial \beta_j}=\sum_{i=1}^{n}\frac{\partial l}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta_j}=0, \tag{8}$$

where:

$$\frac{\partial l}{\partial \theta_i}=\frac{\left(Y_i-\alpha'(\theta_i)\right)}{\varphi}=\frac{Y_i-\mu_i}{\varphi}, \tag{9}$$

$$\frac{\partial \theta_i}{\partial \beta j}=\frac{\partial \theta_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}=\frac{\partial \theta_i}{\partial \eta_i}x_{ij}. \tag{10}$$

Generalized linear models are estimated using Newton-Raphson method, or by method of IRLS (method of iteratively weighted least squares). Using the algorithm Newthon-Rapson can obtain the observed information matrix (OIM), on the contrary, the method of IRLS we obtain the expected information matrix (EIM) see GRAY, Roger J. a Susan M. PITTS (2012).

3.4 Verification

Verification of model parameters is tested by Wald test, model comparison is established by analysis of deviance residuals, Akaike information criterion (AIC) and Bayesian information criterion (BIC).

Wald test

Verification of models parameters is tested by Wald test, model The Wald test is based on a testing hypothesis, it is the way, how to find out if explanatory variables are statistically significant. The null hypothesis:

$$H_0=\hat{\beta}=0, \tag{11}$$

is compared to alternative:

$$H_A=\hat{\beta}\neq 0. \tag{12}$$

If the null hypothesis is rejected, it suggest that parameters should be removed from the model.

Deviance residuals

Using residual analysis is possible to find out the information about the suitability of the model used. This analysis should not be neglected due to wrong selection of type of model, link function, etc. The deviance residuals can be used to assess the quality of the model, e.g. for the detection of remote observation and verification of the assumption about the variance.

Deviance residual is given:

$$r_i^D=sign(y_i-\hat{\mu}_i)\cdot(\sqrt{d(y_i\cdot\mu_i)}, \tag{13}$$

where $(\sqrt{d(y_i\cdot\mu_i)}$ denotes the distance function, which represents the remoteness from the estimated mean values to observed.

Akaike and Bayesian information criteria

The basic of these criteria is the comparison of models among themselves and in the most suitable model is considered to be such a model, the value of which *AIC* and *BIC* is the lowest. Akaike information criterion is following:

$$AIC = 2k - 2\log(L), \tag{14}$$

where k is number of predictors of a model including constants and *log(L)* means log-likelihood model.

Bayesian information criterion is:

$$BIC = -2\log L + k\log(n), \tag{15}$$

n is the number of observations.

4 Estimation the model using GLM

The aim of this paper is to find out ideally suited model for estimation claim frequency based on individual rating factors. For the purposes of this paper it was used a random selection of real data of insurance contracts vehicle insurance concluded on the territory of the Czech republic in 2005-2010, the file contains 18 112 contracts.

All calculations are computed in software STATA 14.0. The estimation includes the following regression characteristics see Table 1. In the framework of the tariff analysis, each category is created using the quartiles. The categories are created by quartiles because of number the observation contained in each category have to be the same.

For the purposes of paper two models are going to be estimated and subsequently these models are going to be compared. First, the full model 1 (called saturated and includes all related variables), which is going to be subsequently compared with the reduced model no. 2. Model no. 2 includes statistically significant variables. Furthermore, these two models are going to be compared to AIC and BIC, and also through the analysis of the deviance residual.

variable	description	category
count	Claim frequency	0,1,2,3
fuel	Type of fuel	1,2,3,4
gender	Gender of driver	0 – man, 1
catagecar	Vehicle age	1 – new car, 2, 3, 4
catageman	Age of driver	1 – the youngest, 2,3,4
catprice	Vehicle price	1 – the cheapest, 2,3,4
catvolumkw	Engine power	1 – the lowest, 2,3,4

Table 1 Data description

There are many options how is possible to find out the influence between dependent (count) and independent variables. For the purpose of the paper is selected the way, where is model estimated with each variable separately. The results are shown in Figure 1.

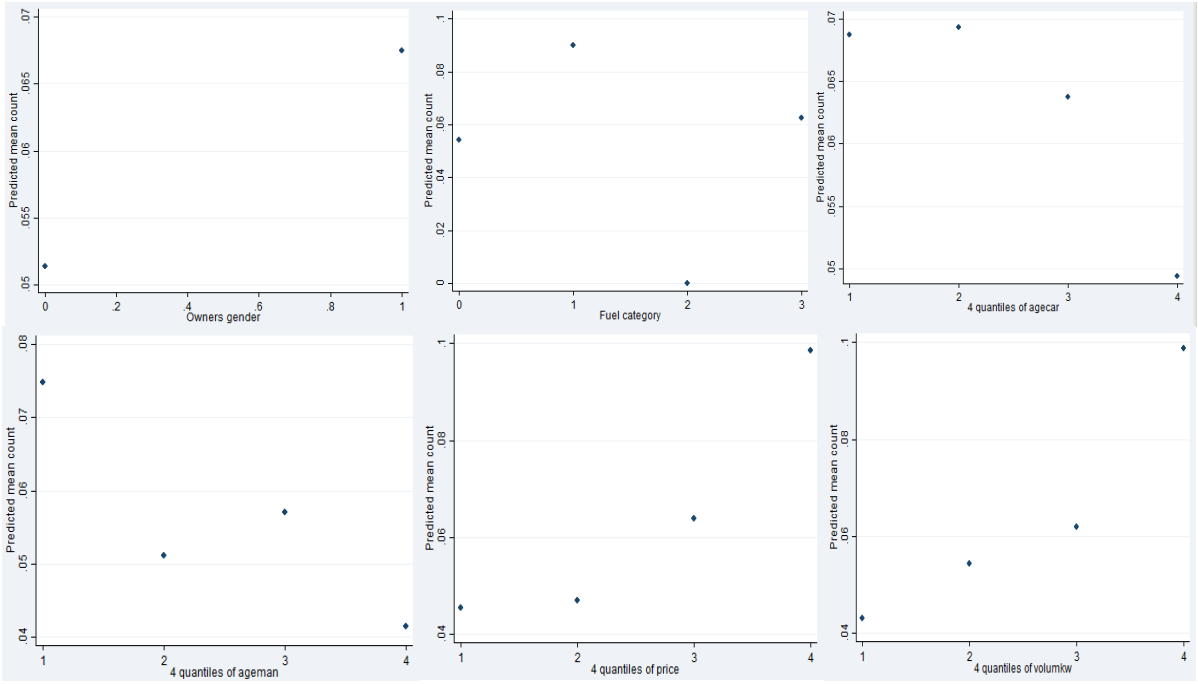


Figure 1 Effect on dependent variable

It should be noted that some of the variables especially *catageman* and *catagecar* were statistically insignificant at confidence level 95 %, just because of this they are subjected by Wald test. In the case of zero hypothesis, see formula (11), is tested the assumption that all the variables of each category separately are equal to zero, that means tested parameter is statistically insignificant, versus alternative hypothesis, see formula (12), that at least one of tested parameter is nonzero. The results of the Wald test see in Table 2.

Variable	**comparison**
$\chi^2(3)$	Significance > χ^2
catageman	
14,77	0,0020
$\chi^2(3)$	Significance > χ^2
catagecar	
42,79	0,0000

Table 2 Wald test

The Wald test is based on the maximum credible estimation. This type of statistic has in large selections division of Pearson $\chi 2$ statistics with the number of freedom, which is equal to difference between the parameters of two tested models. According to the results of Wald test, it is clear that all parameters are statistically significant and should not be excluded from the model.

In the next step the model 1 is reduced by the amount of the statistically insignificant variable, then both models is compared and more appropriate model for the estimation of the claim frequency is selected. The results are shown in Table 3.

variable	**Model 1**	**Model 2**
Deviance residual	0,2683	0,2707
AIC	0,4221	0,4237
BIC	-115032,1000	-115057,2000

Table 3 Comparison

According to the teoretical assumptions, it is clear, that model 2 is for modeling the number of claims more appropriate. In the insurance practice, it is almost impossible to manage the risks for each person separately, because no one has exactly the same input parameters and risk management would be expensive. On the other hand, the tariff groups should be created, where the risk of each group is homogenous.

In the next step, is shown the progress of the residuals of both models, see Figure 2.

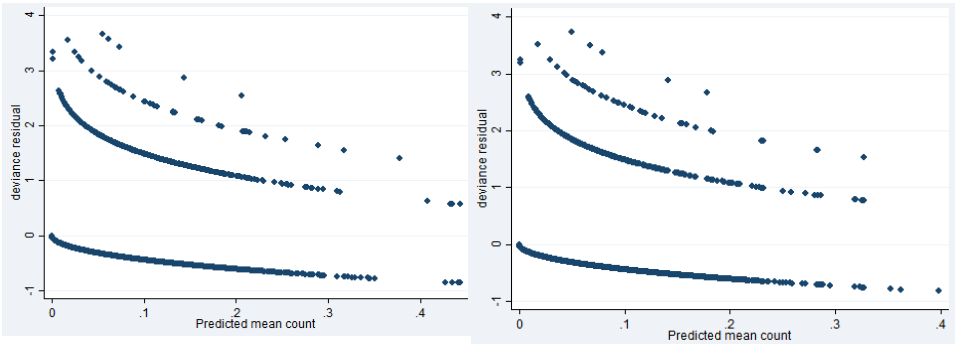


Figure 2 Deviance residuals of model 1 (left) and model 2

According to the results of deviance residuals, it is clear that the model 2 is more appropriate. Model 2 includes less outliers compared to the model 1.

5 Conclusion

The randomness of insurance event and their consequence (i.e., the claim frequency and severity) are a source of uncertainty of the future insurance benefits, which are going to pay out to the clients. This uncertainty can be expressed by an appropriate distribution of probability.

The aim of this paper was to find out ideally suited model for estimation claim frequency based on individual rating factors.

In the first step, it was tested the effect on dependent variable count and then has been selected the appropriate linear regressors for estimation, where some of them looks like statistically insignificant. The parameters of both models were estimated by the maximum likelihood method at standard level of significant 0,05. The probability of distribution was negative-binomial with log-link function.

According to the results of Wald test it has been confirmed, that all involved parameters were statistically significant and should be involve to a model. In the framework of paper were estimated two models. The first, called saturated, and the second reduced and both were subsequently compared by AIC and BIC and Deviance residuals. Estimation was carried out on the categorized data, continuous variables were converted to quartiles. According to the results it can be denoted that second model is more appropriate. A matter of future research could be the tariff analysis.

Acknowledgements

Supported by the SGS Project VŠB-TU Ostrava SP2018/154 „Finanční rozhodování podniků a finančních institucí za rizika“

References

[1] HILBE, Joseph. *Negative binomial regression*. 2nd ed. Cambridge: Cambridge University Press, 2011. ISBN 978-0-521-19815-8.

[2] JONG, Piet de a Gillian Z. HELLER. *Generalized linear models for insurance data*. Cambridge: Cambridge University Press, 2008. ISBN 978-0-521-87914-9.

[3] LONG, J. Scott a Jeremy FREESE. *Regression models for categorical dependent variables using Stata*. Third edition. College Station: Stata Press Publication, 2014. ISBN 978-1-59718-111-2

[4] OHLSSON, Esbjörn a Björn JOHANSSON. *Non-life insurance pricing with generalized linear models*. Berlin: Springer, c2010. ISBN 978-3-642-10790-0.

[5] ŠPAČKOVÁ, Adéla. Estimation of claim frequency by generalized linear models. In: *11th International Conference on Financial Management of Firms and Financial Institutions*. 6.- 7. září. Ostrava: VŠB-TU Ostrava, 2017, s. 821-830. ISBN 978-80-248-4138-0.

Efficiency Assessment of Urban Public Transport Operators: A Case of the Slovak Republic

Eva Štichhauerová[1], Natalie Pelloneová[2]

Abstract. A key aspect in the transportation planning process is, among other things, the efficiency of transportation systems that ensures the required quality of provided services. This paper deals with the efficiency assessment of operators providing public transport services in 27 urban agglomerations in the Slovak Republic in 2016. The objective is to evaluate the performance of selected urban public transport (UPT) operators in terms of technical efficiency using the Data Envelopment Analysis (DEA) method. Two input-oriented DEA models (CCR and BCC) were applied. The number of drivers, the number of lines and the consumption of materials and energy were selected for inputs; the selected outputs were the number of transported persons and sales and revenues from public transport. The application of models led to these results: there were 8 UPT operators identified in Slovakia who are able to transform these inputs into outputs efficiently and supposedly use best practices in the traffic management. Further, 3 UPT operators could be denoted as operating at the optimal and most productive size.

Keywords: Data Envelopment Analysis, efficiency, technical efficiency, urban public transport.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Data Envelopment Analysis is a non-parametric method based on the assumption that, in order to improve performance, it is necessary to measure the efficiency of decision-making units and identify the main sources of inefficiency. Efficiency can be defined as the ratio of the outputs produced by a unit to the inputs consumed by the unit in its operations. This method has become very popular in calculating technical efficiency because it makes it possible to take into account the transformation of multiple inputs into multiple outputs. A decision-making unit can be any unit that produces certain outputs while consuming certain inputs (e.g. banks, schools, public transport operators). A decision-making unit is classified as efficient, if the value of the model's objective function is equal to one. For inefficient units, the value of the purpose function is less/greater than one, depending on the orientation of the selected DEA model. Optimální hodnota účelové funkce je nazývána jako míra technické efektivnosti. Metoda DEA mimo jine The optimal value of the objective function is called technical efficiency. Among other things, the DEA method also provides information on how to improve the behaviour of the assessed unit in order to become efficient [7].

The basic classification of DEA models is according to whether they are input or output oriented. Using input-oriented models, it is possible to calculate how the input indicators of assessed units need to be improved for the units to become efficient. Output-oriented models can be used to calculate how the output indicators of assessed units need to be improved for the units to become efficient. Therefore, in order to achieve technical efficiency we can either focus on reducing inputs (input orientation) or increasing outputs (output orientation) [24]. Another possible criterion to classify DEA models is the assumed type of the returns to scale: constant returns to scale (CCR models) or variable returns to scale (BCC models). For constant returns to scale, the measure of efficiency is called overall technical efficiency, and for variable returns to scale it is called pure technical efficiency [16]. The relevant models are described in more detail in Chapter 2. The DEA method is currently widely used – in professional literature, various studies have been published dealing with performance assessment of public transport systems.

Below is a summary of selected authors who, in their work, apply the DEA method to assess public transport performance and efficiency. They are listed in chronological order by year of publication and subsequently alphabetically by author. Agarwal [1] measures the technical efficiency of public transport in India. In his paper, he provides an overview of the situation of state-owned transport companies in terms of their production efficiency for 2004–2005 and 2007–2008. Efficiency is measured using a DEA model. The following inputs are

[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

considered: fleet size, total number of employees and fuel consumption. The number of passenger-kilometres is selected as the output. García-Sánchez [9] uses the DEA method to perform a comparative analysis of the efficiency of public bus transport in Spain. The completed research of Spanish public transport has shown that the pure technical efficiency averages 94.91% and scale efficiency 52.02%. Hilmola [13] assesses public transport efficiency in 52 major cities both within and outside Europe on data from 2001. To assess efficiency, he uses four different comparative-analysis models that are based on the DEA method. Five inputs and five outputs were used in the efficiency measurement model. Pamučar and Boban [18] show that the DEA method can be applied to the quality management of public passenger transport networks. They use the DEA method to determine the optimal quality parameters, thus helping to define the total cost of public passenger transport lines, or the total amount of funds needed for the normal operation of the public transport network. Caulfield et al. [5] addresses the issue of providing fast, frequent and modern public transport traffic from the centre of Dublin to the local airport. The main objective is to explore and identify the most efficient transport route solutions: the centre of Dublin – the airport. They use the DEA method to identify the most efficient route and to subsequently determine the causes of inefficiency of the other routes. Higashimoto et al. [12] used the DEA method to carry out an analysis of the efficiency of the network of urban bus routes in Tomakomai, Japan, aiming to structure bus routes in terms of optimising route layouts and attracting customers. The following parameters were selected for assessing the efficiency of bus transport: route length, number of bus stops, population along the line, transportation cost and number of passengers. Kráľ and Roháčová [15] use a combination of DEA methods and stochastic frontier analysis (SFA) to measure and compare the efficiency of public road transport companies in the Slovak Republic. For this purpose, they use an input-oriented DEA model that assumes variable returns to scale. Li et al. [17] present the use of the DEA method to assess the performance of bus routes within the public transport system and, in turn, attempt to improve the overall operation of the selected public transport system. The selected DEA method is applied to three bus routes within Beijing's public transport system. Fancello et al. [8] deal with improving the effectiveness of transport networks and improving the performance of the road system. In their paper they use the DEA method to compare the performance of different urban networks in order to provide policy-makers with technical support to correctly select the actions to be taken to improve the urban road system. Vaidya [22] assesses the relative performance of 26 urban transport organisations in India using the CCR DEA approach. The assessment relies on a total of 19 different criteria. Ayadi and Hammami [2] study the cost effectiveness of the Tunisian public transport industry using the DEA method. Their study is carried out based on 12 samples of regional transport companies for the period 2000–2010. Carvalho et al. [4] assess urban public transport (UPT) in the twenty largest Brazilian cities through an output-oriented DEA model. The paper aims to compare selected cities using collected secondary data from 2005–2010 in terms of three performance measures: infrastructure efficiency, service efficiency, and efficiency score. In their paper, they chose the following variables: number of inhabitants of the municipality, number of buses, average daily number of passengers, and average price. Hahn et al. [11] used the DEA model to assess the sustainability of public transport that focuses on fast bus routes in Seoul. The resulting model reflects an interaction of three characteristics of sustainable transport services, namely efficiency, equality and environmental impact. Sow et al. [20] explore transport systems efficiency, which they consider to be the most important step within the transport planning process. Their paper focuses on assessing the efficiency of the routes of the Dakar Dem Dikk transport company that operates in Dakar, Senegal. Their research aims to use the DEA method in order to identify major opportunities for improvement. The paper examines the efficiency of 24 lines using two output-oriented DEA models, i.e. both the CRS model and the VRS model. Tran et al. [21] deal with the measurement of the performance of routes within a transit system. The DEA method was used to compare the performance of transit operators and transit lines routes within a transit system. In a case study on a bus route in Australian Brisbane, temporal performance evaluation was assessed. Number of service and total travel time were selected as inputs. Transit work and on-time performance were selected as outputs.

2 Data and Methodology

The main source of data was the Annual Report of the Association of public passenger transport operators in urban agglomerations in the Slovak Republic [23]. The Association of public passenger transport operators is an association of legal entities whose members include Slovakia's 27 most important UPT operators. The Annual Report for 2016 contains detailed information on the vehicle fleet, investments, the number of employees, the number of lines, revenues, costs, transport capacities etc. From the agglomerations perspective, bus services are operated in all agglomerations. Trolleybus services are operated in 4 agglomerations (Bratislava, Žilina, Prešov, Banská Bystrica, while in Košice trolleybus services were suspended in 2015), tram services are operated in 2 agglomerations (Bratislava, Košice). With the exception of the transport company of Banská Bystrica, all of the above operators operate bus services. Only 1 operator – the transport company of Bratislava – provides transport using all types of means of transport.

Based on available data from the above source, it was also necessary to select appropriate sets of inputs (I) and outputs (O). The following inputs were considered: I1 – Total number of vehicles, where more detailed information on the bus, trolleybus and tram fleets is also available. Since public transport operators are assessed regardless of the type of transport, such detailed information cannot be used. I2 – Investments in transport vehicles (EUR). I3 – Average number of employees (FTE). I4 – Number of drivers, where more detailed information on bus, trolleybus and tram drivers is also available but not used here. I5 – Wages and salaries (EUR). I6 – Number of lines operated. I7 – Length of lines operated (km). I8 – Tangible and intangible assets at purchase price (EUR). I9 – Tangible and intangible assets at purchase price, of which transport vehicles (EUR). I10 – Total costs (EUR). I11 – Consumption of materials and energy (EUR). I12 – Depreciation (EUR). The following outputs were considered: O1 – Total transport capacity (vehicle-kilometres). O2 – Number of passengers transported. O3 – Revenues from passengers, including fines (EUR). O4 – Total revenues (EUR). O5 – Revenues from UPT, including fines (EUR).

Another important step is to choose the appropriate number of inputs and outputs, taking into account the number of DMUs. While Golany and Roll [10] set the rule that the number of DMUs should be at least twice as high as the number of the inputs and outputs considered, Bowlin [3] requires a number that is three times as high, and Cooper et al. [6] even recommend that the number of DMUs should be at least equal to the sum of inputs and outputs. The authors of this paper favour the stricter limit according to Bowlin [3], because they share the view that a high number of inputs and outputs relative to the number of DMUs would negatively affect the information value of the DEA method. In the first step, variables for which complete data were not available for all UPT operators (i.e. I7, I8 and I9) were excluded from the selection. The number of inputs and outputs can be further reduced using correlation analysis. One of the main requirements for selecting variables is a relatively high between-group correlation, i.e. all outputs should be generated by the inputs. For this purpose, Spearman's correlation coefficient was calculated for all variables. Table 1 shows the correlation matrix of the variables considered. Based on an assessment of the matrix, the following 3 inputs were used: (I4) number of drivers, (I6) number of lines operated, and (I11) consumption of materials and energy (EUR), while the 2 outputs used were: (O2) number of passengers transported, and (O5) revenues from UPT, including fines (EUR).

	I1	**I2**	**I3**	**I4**	**I5**	**I6**	**I10**	**I11**	**I12**	**O1**	**O2**	**O3**	**O4**	**O5**
I1	x													
I2	0.328	x												
I3	0.741	0.284	x											
I4	0.910	0.302	0.876	x										
I5	0.946	0.323	0.801	0.918	x									
I6	0.888	0.466	0.648	0.830	0.846	x								
I10	0.916	0.433	0.709	0.857	0.909	0.896	x							
I11	0.897	0.423	0.750	0.860	0.931	0.800	0.888	x						
I12	0.939	0.293	0.712	0.855	0.900	0.795	0.891	0.863	x					
O1	0.959	0.362	0.757	0.918	0.952	0.913	0.932	0.937	0.910	x				
O2	0.889	0.250	0.688	0.845	0.951	0.821	0.854	0.855	0.852	0.915	x			
O3	0.959	0.271	0.792	0.914	0.929	0.890	0.907	0.858	0.921	0.957	0.896	x		
O4	0.969	0.322	0.777	0.921	0.959	0.876	0.949	0.932	0.950	0.974	0.904	0.966	x	
O5	0.962	0.311	0.787	0.919	0.931	0.899	0.926	0.866	0.917	0.965	0.887	0.994	0.967	x

Table 1 Correlation coefficients among input and output variables

The next step is to construct DEA models and determine the technical efficiency (TE) score. In measuring the efficiency of 27 UPT operators in the Slovak Republic using the DEA method, input-oriented DEA models were used, namely the BCC-I and CCR-I models. The classical BCC-I model assumes variable returns to scale and can be written as equation (1) while the condition of convexity (2) is met. In this equation, λ_j are weights of all DMUs, s^-_i and s^+_r are slack variables, $\varepsilon > 0$ is an infinitesimal constant (so-called non-Archimedean element) defined to be smaller than any positive real number, and θ is the efficiency score that expresses the reduction rate of inputs in order this unit reaches the efficient frontier. The BCC model measures pure technical efficiency (PTE). In the case of the CCR-I model operating under constant returns to scale, the condition of convexity (2) is left out of the model. The CCR model measures overall technical efficiency (OTE), by aggregating pure technical efficiency and scale efficiency into a single value. In addition, the TE score was determined for each operator. All calculations were made using the OSDEA-GUI software. A score of 1 is obtained by an operator that does not show signs of inefficiency when compared with other related operators. A score of less than 1 indicates an inefficient operator.

$$E_0 = \min .\theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right)$$

$$s.t. \sum_{j=1}^{n} \lambda_j X_{ij} + s_i^- = \theta X_{i0}, i = 1,...,m, \qquad (1)$$

$$\sum_{j=1}^{n} \lambda_j Y_{rj} - s_i^+ = Y_{r0}, r = 1,...,s,$$

$$\lambda_j, s_i^-, s_i^+ \geq 0, j = 1,...,n, i = 1,...,m, r = 1,...,s. \ \theta \text{ unrestricted in sign.}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \qquad (2)$$

The relationship between the three measures of efficiency is expressed by equation (3). The scale efficiency measures the degree to which an enterprise can improve its efficiency by changing its size. The scale efficiency score is determined according to equation (3).

$$SE = \frac{OTE}{PTE} \qquad (3)$$

The source of scale inefficiency may be either decreasing returns to scale (DRS) or increasing returns to scale (IRS). The specific type of returns to scale (RTS) can be determined using the sum of the weights λ of peer units [24]. If $\sum\lambda_j = 1$, the operator is scale efficient and operates under constant returns to scale. If $\sum\lambda_j > 1$ the operator operates under DRS conditions and inefficiency can be addressed through reducing the scale of operations. If $\sum\lambda_j < 1$ the operator exists under IRS conditions and it can expand the scale of its operations.

3 Research results

First, the BCC-I model was applied to the data. This model is based on the assumption of variable returns to scale and provides the pure technical efficiency (PTE) score. The BCC-I model classified 8 (i.e. 30 %) operators as efficient, i.e. as being able to efficiently transform the inputs into the outputs. It can be assumed that transport companies use appropriate procedures and processes in their operations. According to the PTE score, a ranking was made where all efficient operators were ranked in the first position. The remaining transport companies, which were classified as inefficient according to the BCC-I model, require an amount of inputs that is much greater than optimal in order to generate the given level of outputs. Their position in the ranking was 9th–27th; the worst ranked operator was Michalovce - DZS – M.K. TRANS. In an input-oriented model, the general recommendations for such inefficient operators include changing the practices and processes in their operations so as to reduce the level of inputs at the current level of outputs. The average PTE score is 0.7333. Detailed results are shown in Table 2.

Subsequently, the overall technical efficiency (OTE) score was determined using the CCR-I model. The CCR-I model classified 3 UPT operators (i.e. 11 %) as efficient – they operate at the optimal scale, i.e. the size of their operations is optimal and most productive. The average OTE score is 0.5549. Detailed results are shown in Table 2.

The table also shows that all operators that were classified as efficient according to the CCR-I model are also classified as efficient according to the BCC-I model. This means that the size of UPT operators' operations is optimal and, at the same time, that these operators are able to efficiently transform the given inputs into outputs thanks to appropriate management methods, practices and processes.

It is clear that the PTE score in the BCC-I model is higher than the OTE score in the CCR-I model, because the CCR-I model takes into account scale inefficiency. The scale efficiency (SE) scores calculated according to equation (3) are shown in Table 2. The average SE score is 0.7713 and the maximum SE value is 1 (the maximum value is achieved by 3 operators that are classified as efficient according to the CCR-I model).

For the 5 transport companies that are classified as inefficient according to the CCR-I model yet as efficient according to the BCC-I model, it can be concluded that their technical inefficiency is caused by scale inefficiency. It can be assumed that, while these transport companies use best practices in operations management, their size is not optimal. The management recommendations for these selected operators should focus on changing the scale of operations (i.e. size) depending on the type of returns to scale.

Operator	OTE	PTE	SE	∑λ	RTS	Ranking
Bratislava	**1.0000**	**1.0000**	**1.0000**	1.0000	CRT	1st
Košice	0.7626	0.7631	0.9993	0.9488	IRS	13th
Žilina	0.8088	0.8394	0.9635	0.3420	IRS	11th
Prešov	0.6704	0.6737	0.9951	0.8641	IRS	15th
Banská Bystrica	0.7914	0.9177	0.8624	0.1558	IRS	10th
Považská Bystrica	0.4167	0.4780	0.8718	0.2169	IRS	24th
Púchov	0.2445	0.5041	0.4850	0.0304	IRS	23rd
Trnava	**1.0000**	**1.0000**	**1.0000**	1.0000	CRT	1st
Trenčín	0.5179	0.5720	0.9054	0.2799	IRS	18th
Prievidza	0.3904	0.4498	0.8679	0.2584	IRS	25th
Zvolen	0.5068	0.5769	0.8785	0.2349	IRS	17th
Žiar nad Hronom	0.4294	**1.0000**	0.4294	0.0238	IRS	1st
Banská Bystrica - SAD	0.6264	0.6595	0.9498	0.3766	IRS	16th
Brezno	0.7265	**1.0000**	0.7265	0.0501	IRS	1st
Nové Zámky	0.4702	0.5208	0.9028	0.1550	IRS	22nd
Ružomberok - Arriva Liorbus	0.5046	0.5323	0.9480	0.2673	IRS	20th
Martin	0.7775	0.7910	0.9829	0.7035	IRS	12th
Košice - Eurobus	0.3766	0.4170	0.9031	0.1575	IRS	26th
Poprad	**1.0000**	**1.0000**	**1.0000**	1.0000	CRT	1st
Nitra	0.7120	0.7265	0.9800	0.5195	IRS	14th
Lučenec	0.4393	0.5526	0.7950	0.0915	IRS	19th
Bardejov	0.8095	0.9533	0.8492	0.1013	IRS	9th
Skalica	0.3309	**1.0000**	0.3309	0.0102	IRS	1st
Michalovce - DZS – M.K. TRANS	0.2265	0.3413	0.6636	0.0465	IRS	27th
Humenné - DZS – M.K. TRANS	0.1045	0.5305	0.1970	0.0152	IRS	21st
Stara Ľubovňa - DZS – M.K. TRANS	0.1690	**1.0000**	0.1690	0.0098	IRS	1st
Michalovce - Arriva	0.1688	**1.0000**	0.1688	0.0094	IRS	1st
Average efficiency	**0.5549**	**0.7333**	**0.7713**			
SD	**0.2597**	**0.2226**	**0.2727**			
MIN	**0.1045**	**0.3413**	**0.1688**			
MAX	**1.0000**	**1.0000**	**1.0000**			

Table 2 Comparison of efficiency scores of operators

Table 2 also indicates the type of returns to scale for all 27 UPT operators, which is derived based on the sum of the weights λ of peer units. It is clear that only 3 carriers operate at the optimal scale, namely the carriers that operate UPT in Bratislava, Trnava and Poprad – these operate under constant returns to scale. Due to increasing returns to scale, the remaining 24 UPT operators can be considered smaller that their optimal scale size. From this perspective, they could resolve their technical inefficiency through expanding their operations, and it can be expected that an increase in inputs will generate a more than proportional increase in outputs.

4 Conclusion

The paper presents the results of research on the technical efficiency assessment of 27 UPT operators in the Slovak Republic. The results of the research may serve as a methodological proposal for assessing the performance of UPT operators. DEA was selected as the assessment method, namely the input-oriented BCC-I and CCR-I models in which the technical efficiency of the transformation of 3 inputs into 2 outputs was measured. The average OTE score was 0.5549, the average PTE score was 0.7333. The average scale efficiency of Slovak UPT operators was determined at 0.7713. In productivity terms, the average UPT operator is smaller than its optimal scale size because it operates under increasing returns to scale.

By applying two input-oriented DEA models, a ranking was compiled and 8 UPT operators (i.e. 30 %) were identified that are able to efficiently transform the given inputs into outputs, as they use best practices and appropriate processes in operations management. Of these UPT operators, 3 (i.e. 11% of all 27) could also be classified as UPT operators that achieve the optimal and most productive size, because they operate under constant returns to scale. Furthermore, 5 UPT operators were identified that likely use best practices in operations management and are able to efficiently transform inputs into outputs, yet their size is smaller than optimal. In such cases, it is advisable to consider expanding operations.

A total of 19 UPT operators were classified as inefficient in both of the aspects under review, i.e. both in terms of the efficiency of the transformation of inputs into outputs and in terms of size. From management

perspective, it can be recommended that the focus should be primarily on improving the efficiency of the transformation process. UPT operators should adopt best practices. At the same time, the optimal size of operations should also be taken into consideration in response to the fact that these UPT operators currently operate under increasing returns to scale.

The impacts of applying the selected DEA models could be expanded to include determining the peer units for each of the inefficient UPT operators, which would represent the absolute best practice. Each peer unit is characterised by its weight, which can be used to determine the level of inputs required in order to make a thus far inefficient UPT operator reach the efficient frontier. Another opportunity for further research could be employing an appropriate ranking method for a more precise evaluation of the scale efficient UPT operators.

References

[1] Agarwal, S. (2009). Measuring the Efficiency of Public Transport Sector in India: An Application of Data Envelopment Analysis. In *Proceedings of a Conference in Philadelphia*, (pp. 1–19). Philadelphia.

[2] Ayadi, A. & Hammami, S. (2015). Efficiency of public transport system in Tunisia: DEA–SFA approach. *International Journal of Economics and Business Research, 9(3)*, 316–336.

[3] Bowlin, W. F. (1998). Measuring Performance: An Introduction to Data Envelopment Analysis (DEA). *Journal of Cost Analysis, 7*, 3–27.

[4] Carvalho, M., et al. (2015). Efficiency and Effectiveness Analysis of Public Transport of Brazilian Cities. *Journal of Transport Literature, 9(3)*, 40–44.

[5] Caulfield, B., et al. (2013). Using data envelopment analysis as a public transport project appraisal tool. *Transport Policy, 29*, 74–85.

[6] Cooper, W. W., et al. (2000). *A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Norwell: Kluwer Academic Publishers.

[7] Cooper, W. W., et al. (2011). *Handbook on Data Envelopment Analysis*. New York: Springer Science and Business Media.

[8] Fancello, G., et al. (2014). Data Envelopment Analysis (D.E.A.) for urban road system performance assessment. *Procedia - Social and Behavioral Sciences, 111*, 780–789.

[9] García-Sánchez, I. M. (2009). Technical and Scale Efficiency in Spanish Urban Transport: Estimating with Data Envelopment Analysis. *Advances in Operations Research, 2009*, 1–15.

[10] Golany, B. & Roll, Y. (1989). An Application Procedure for DEA. *Omega, 17*(3), 237–250.

[11] Hahn, J., et al. (2017). Sustainability evaluation of rapid routes for buses with a network DEA model. *International Journal of Sustainable Transportation, 11*(9), 659–669.

[12] Higashimoto, Y., et al. (2013). Evaluation of bus route efficiency by network DEA including social priority. *Journal of the Eastern Asia Society for Transportation Studies, 10*, 1323–1333.

[13] Hilmola, O. (2011). Benchmarking efficiency of public passenger transport in larger cities. *Benchmarking: An International Journal, 18*(1), 23–41.

[14] Ji, Y. & Lee, C. (2010). Data Envelopment analysis. *The Stata Journal, 10*(2), 267–280.

[15] Kráľ, P. & Roháčová, V. (2013). Measuring the Efficiency of Public Road Transport Companies in the Slovak Republic Using DEA and SFA. *Statistika, 93*(2), 76–85.

[16] Kumar, S. & Gulati, R. (2008). An Examination of Technical, Pure Technical, and Scale Efficiencies in Indian Public Sector Banks using Data Envelopment Analysis. *Eurasian Journal of Business and Economics, 1*(2), 33–69.

[17] Li, J., et al. (2013). Evaluation of Public Transportation Operation Based on Data Envelopment Analysis. *Procedia - Social and Behavioral Sciences, 96*, 148–155.

[18] Pamučar, S. D. & Boban, D. D. (2012). DEA model for measuring the quality control of public transport networks. *Vojnotehnički Glasnik/Military Technical Courier, 60*(2), 28–47.

[19] Roháčová, V. (2015). A DEA based approach for optimization of urban public transport system. *Central European Journal of Operations Research, 23*(1), 215–233.

[20] Sow, O., et al. (2016). Efficiency Analysis of Public Transportation Subunits Using DEA and Bootstrap Approaches – Dakar Dem Dikk Case Study. *Journal of Mathematics Research, 9*(6), 114–127.

[21] Tran, K., et al. (2016). Data envelopment analysis (DEA) based transit route temporal performance assessment: a pilot study. In *38th Australasian Transport Research Forum* (pp. 1–18). Melbourne.

[22] Vaidya, O. S. (2014). Evaluating the Performance of Public Urban Transportation Systems in India. *Journal of Public Transportation, 17*(4), 174–191.

[23] Združenie prevádzkovateľov hromadnej dopravy osôb v mestských aglomeráciách Slovenskej republiky. (2018). *Ročenka združenie prevádzkovateľov hromadnej dopravy osôb v mestských aglomeráciách Slovenskej republiky 2016*, Bratislava.

[24] Zhu, J. (2014). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets*. Cham: Springer International Publishing.

VEHICLE ROUTING WITH THREE TYPES OF COST FUNCTIONS: A RANDOMIZED HEURISTIC APPROACH

Petr Štourač[1], Michal Černý[2], Miroslav Rada[3]

Abstract. We consider a version of Vehicle Routing with three types of carriers using different forms of cost functions. The costs of delivery of Carrier I are proportional to the sum of distances traveled by the vehicles in his fleet. The costs of Carrier II are proportional to the distance traveled and the weight transported. For Carrier III, the costs are proportional to the weight transported only. This conference talk builds on our recent working paper addressing this problem, where two heuristic algorithms were designed. The analysis of computational results lead us to further ideas for improvement, some of which are summarized in this talk. Namely, we can treat better the loading of vehicles of Carrier I and we also implement a randomized step which can, in some cases, avoid locally poor solutions given by the deterministic heuristic algorithms. We present a computational study and compare the performance of the original and improved algorithms.

Keywords: Vehicle routing, binary integer programming, heuristic.

JEL classification: C61
AMS classification: 90C59

1 Introduction

A combination method of transporting batches is one of the many variants of Vehicle Routing Problem (other possibilities are presented in [1], [4]). This method can be further divided into the two basic types. In the first one, the batches distributor has more carriers with combined types of fleet and different transport costs, and decides which carrier to deliver the particular shipment. An example of this type can be a vendor with a central depot, which ensures delivery of batches to its customers. In the second case, the distributor deals with the problem of transporting one batches with the help of multiple carriers with different costs. This option is typical for the transport of batches across the sea by means of combined transport by vehicle and by aircraft or vehicle and by boat. Another possibility may be transporting of batches by means of a combination of two basic types.

This paper deals with a simple variant of the combined transport of the first type shipments with three carriers and builds on our previous working paper [6], where the mathematical model and two heuristics were defined. The analysis of computational results lead us to further ideas for improvement. Namely, we can treat better the loading of vehicles of Carrier I and we also implement a randomized step which can, in some cases, avoid locally poor solutions given by the deterministic heuristic algorithms.

We consider a version of Vehicle Routing with three types of carriers using different forms of cost functions. The costs of delivery of Carrier I are proportional to the sum of distances traveled by the vehicles in his fleet. The costs of Carrier II are proportional to the distance traveled and the weight transported. For Carrier III, the costs are proportional to the weight transported only. We define three new heuristics. These ones again assume the same behavior for Carriers I and II as TSP style (heuristic algorithms for the Traveling Salesman Problem are described in [5], [3]) and use insert heuristics to optimize the route like in [2]. Heuristics III modifies Heuristic II by adding randomized step for choosing the cheapest variant in every step of processing. The algorithm in each new state of processing calculates for individual carriers and their cars the cheapest option to deliver other packages and from these variants selects according the random value the first or the second cheapest one. Heuristics IV modifies Heuristic II by adding algorithm for finding better solution of vehicle loading of Carrier I vehicles. After finishing of calculation, if some vehicles of Carrier I are not completely loaded, the algorithm tries to find better solution for vehicle loading. Heuristics V joins algorithm of previous Heuristics III and Heuristic IV.

[1]University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, nm. W. Churchilla 1938/4, Prague, Czech Republic, stop02@vse.cz
[2]University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, nm. W. Churchilla 1938/4, Prague, Czech Republic, cernym@vse.cz
[3]University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, nm. W. Churchilla 1938/4, Prague, Czech Republic, miroslav.rada@vse.cz

2 ILP model

As in the paper [6], a mathematical model was used to compare heuristics performance. The definition of objective function (1) with the objective to minimize the sum of cost functions of the three carriers and constraints (2) . . . (12) is repeated here.

Let $G = \{V, E\}$ be a complete undirected graph with $V = \{1, 2, \ldots, n\}$, where node 1 is the depot.

Input data:

- $m =$ number of batches,
- $\ell =$ number of vehicles of Carrier I,
- $d_{ij} =$distance between nodes i, j (generally, the matrix (d_{ij}) need not be symmetric),
- $w_k =$ weight of batch k,
- $c_q =$ capacity of the q-th vehicle of Carrier I,
- $p_q =$ costs per km of the q-th vehicle of Carrier I,
- $p_{\text{Carr\_II}} =$ costs per km of the vehicle of Carrier II,
- $r_{\text{Carr\_II}} =$ costs per ton loaded by Carrier II,
- $r_{\text{Carr\_III}} =$ costs per ton transported by Carrier III,
- $z_{ki} \in \{0, 1\}$, $z_{ki} = 1$ iff node i is the destination for batch k.

Model variables:

- $y_k^q \in \{0, 1\}$, $y_k^q = 1$ iff batch k is transported by vehicle q of Carrier I,
- $x_{i,j}^q \in \{0, 1\}$, $x_k^q = 1$ iff vehicle q of Carrier I travels from node i to node j,
- $y_k^{\mathbf{s}} \in \{0, 1\}$, $y_k^{\mathbf{s}} = 1$ iff batch k is transported by Carrier II,
- $x_{i,j}^{\mathbf{s}} \in \{0, 1\}$, $x_{i,j}^{\mathbf{s}} = 1$ iff the vehicle of Carrier II travels from node i to node j,
- $y_k^{\mathbf{t}} \in \{0, 1\}$, $y_k^t = 1$ iff batch k is transported by Carrier III,
- u_i^q are auxiliary variables in anti-cyclic constraints for Carrier I,
- $u_i^{\mathbf{s}}$ are auxiliary variables in anti-cyclic constraints for Carrier II.

Objective function:

$$\min \sum_{q,i,j} p_q d_{ij}^q x_{ij}^q + p_{\text{Carr\_II}} \sum_{i,j} d_{ij}^{\mathbf{s}} x_{ij}^{\mathbf{s}} + r_{\text{Carr\_II}} \sum_k w_k y_k^{\mathbf{s}} + r_{\text{Carr\_III}} \sum_k w_k y_k^{\mathbf{t}}. \tag{1}$$

Constraints:

$$\sum_i x_{ij}^q = \sum_i x_{ji}^q \qquad \forall q, j, \tag{2}$$

$$\sum_i x_{ij}^{\mathbf{s}} = \sum_i x_{ji}^{\mathbf{s}} \qquad \forall j, \tag{3}$$

$$\sum_k w_k y_k^q \le c_q \qquad \forall q, \tag{4}$$

$$\sum_q y_k^q + y_k^{\mathbf{s}} + y_k^{\mathbf{t}} = 1 \qquad \forall k, \tag{5}$$

$$\sum_{j, j \ne i} x_{ji}^q \le \sum_k y_k^q z_{ki} \qquad \forall q, \forall i \ge 2, \tag{6}$$

$$\sum_{j, j \ne i} x_{ji}^{\mathbf{s}} \le \sum_k y_k^{\mathbf{s}} z_{ki} \qquad \forall i \ge 2, \tag{7}$$

$$\sum_{j,k, j \ne i} z_{ki} x_{ji}^q \ge \sum_k y_k^q z_{ki} \qquad \forall q, \forall i \ge 2, \tag{8}$$

$$\sum_{j,k, j \ne i} z_{ki} x_{ji}^{\mathbf{s}} \ge \sum_k y_k^{\mathbf{s}} z_{ki} \qquad \forall i \ge 2, \tag{9}$$

$$\sum_j x_{1j}^q \le 1 \qquad \forall q, \tag{10}$$

$$u_i^q - u_j^q + n x_{ij}^q \le n - 1 \qquad \forall q, i, \forall j \ge 2, i \ne j, \tag{11}$$

$$u_i^s - u_j^{\mathbf{s}} + n x_{ij}^{\mathbf{s}} \le n - 1 \qquad \forall i, \forall j \ge 2, i \ne j. \tag{12}$$

3 Heuristics

The paper [6] describes two heuristics for solving the problem, which at each processing step are looking for the cheapest variant to deliver one of the packages that have not yet been delivered. The difference in the approach of both heuristics is that Heuristic I delivers batches in the order they were entered at the input, while Heuristic II chooses the cheapest delivery variant from all packages. After assigning the selected batch to the vehicle, the following points are applied:

- if the batch is assigned to Carrier I, other batches into the same node are loaded if possible (used in Heuristic 1 and Heuristic II),
- if the batch is assigned to Carrier II, additional batches for the same node are handled on the condition that the delivery will be cheaper than for Carrier III (used in Heuristic II).

In this section, we design three new heuristics for solving the problem, which are based on the algorithm of Heuristic II. Heuristics III adds a randomized step for choosing the cheapest variant in every step of processing. The algorithm in each new state of processing calculates for individual carriers and their cars the cheapest option to deliver other packages and from these variants selects according the random value the first or the second cheapest one. Heuristics IV adds algorithm for finding a better solution of vehicle loading of Carrier I vehicles. After finishing of calculation, if some vehicles of Carrier I are not completely loaded, the algorithm tries to find a better solution for vehicle loading. Heuristics V joins algorithm of both previous Heuristics III and Heuristic IV.

For the description of the heuristics, we use the same notation as was described in [6].

3.1 Heuristic III

Heuristic III modifies Heuristic II by adding the randomized step for choosing the cheapest variant in every step of processing. The algorithm in each new state of processing (current carrier position) calculates, for individual carriers and their cars, the cheapest option to deliver other packages (i.e. to which node they will go and what they will carry) and from these variants selects according the random value the first or the second cheapest one. The algorithm ends after processing the last batch. The algorithm of Heuristics III is calculated repeatedly (e.g. 10 000 times) for finding the best solution.

Step 1: Initialize start parameters. All sequences are initialized by value $(1, 1)$, i.e start and stop node must be City 1 (depot).

Step 2: Calculate delivery costs for every vehicle. The costs for each vehicle and each undelivered batch are calculated as follows:

- $\mu_k^q = \infty$if $w_k > c_q - \sum_{\ell, \ell \in Y^q} w_\ell$, else $\mu_k^q = \min_{a_j}\{p_q(d_{a_j,i_k} + d_{i_k,a_{j+1}} - d_{a_j,a_{j+1}})\}$ where $a_j \in (a_j^q) \wedge j < N_q$, $i_k =$ the destination node for batch k,
- $\nu_k = \min_{b_j}\{p_{\text{Carr\_II}}(d_{b_j,i_k} + d_{i_k,b_{j+1}} - d_{b_j,b_{j+1}}) + w_k r_{\text{Carr\_II}}\}$ where $b_j \in (b_j) \wedge j < M$, $i_k =$ the destination node for batch k,
- $\omega_k = w_k r_{\text{Carr\_III}}$.

For each vehicle minimal costs are selected over all undelivered batches:

- $\mu_{k_I}^q = \min_k\{\mu_k^q\}$ where the decision on the minimum is made on the basis of expression μ_k^q / w_k,
- $\nu_{k_{II}} = \min_k\{\nu_k\}$,
- $\omega_{k_{III}} = \min_k\{\omega_k\}$.

Step 3: Choose the carrier's vehicle with minimal costs by using randomized step. The costs are calculated as follows:

Find minimal costs and the second one

- $\sigma_1 = \min_q\{\mu_{k_I}^q, \nu_{k_{II}}, \omega_{k_{III}}\}$.
- $\sigma_2 = \min_q\{\{\mu_{k_I}^q, \nu_{k_{II}}, \omega_{k_{III}}\} - \{\sigma_1\}\}$.

Calculate costs with randomized step

- $\sigma = rnd\{\sigma_1, \sigma_2\}$.

For the minimizing carrier and vehicle, do the following:

Carrier I: add additional undelivered batches to the vehicle for the destination (if it can carry), add another node to the vehicle route (a_j^q), mark laden batches as delivered. Update the set of delivered batches Y^q for the vehicle q with laden batches.

Carrier II: add additional undelivered batches to the destination on the condition that their delivery is cheaper than if they were delivered by Carrier III, add another node to the vehicle route (b_j), mark laden batch as delivered. Update the set of delivered batches $Y^{\mathbf{s}}$ with laden batches.

Carrier III: mark laden batch as delivered. Update the set of delivered batches $Y^{\mathbf{t}}$ with laden batches.

If there is still an undelivered batch, take the first one and proceed to Step 2, otherwise the processing will stop.

3.2 Heuristic IV

Heuristics IV modifies Heuristic II by adding algorithm for finding a better solution of vehicle loading of Carrier I vehicles. The algorithm in each new state of processing (current carrier position) calculates, for individual carriers and their cars, the cheapest option to deliver other packages (i.e. to which node they will go and what they will carry). After finishing of the calculation, if some vehicles of Carrier I are not completely loaded, the algorithm tries to find a better solution for vehicle loading.

Step 1: Initialize start parameters. All sequences are initialized by value $(1, 1)$, i.e start and stop node must be City 1 (depot).

Step 2: Calculate delivery costs for every vehicle. The costs for each vehicle and each undelivered batch are calculated as follows:

- $\mu_k^q = \infty$if $w_k > c_q - \sum_{\ell, \ell \in Y^q} w_\ell$, else $\mu_k^q = \min_{a_j}\{p_q(d_{a_j,i_k} + d_{i_k,a_{j+1}} - d_{a_j,a_{j+1}})\}$ where $a_j \in (a_j^q) \wedge j < N_q$, $i_k =$ the destination node for batch k,
- $\nu_k = \min_{b_j}\{p_{\text{Carr\_II}}(d_{b_j,i_k} + d_{i_k,b_{j+1}} - d_{b_j,b_{j+1}}) + w_k r_{\text{Carr\_II}}\}$ where $b_j \in (b_j) \wedge j < M$, $i_k =$ the destination node for batch k,
- $\omega_k = w_k r_{\text{Carr\_III}}$.

For each vehicle minimal costs are selected over all undelivered batches:

- $\mu_{k_I}^q = \min_k\{\mu_k^q\}$ where the decision on the minimum is made on the basis of expression μ_k^q / w_k,
- $\nu_{k_{II}} = \min_k\{\nu_k\}$,
- $\omega_{k_{III}} = \min_k\{\omega_k\}$.

Step 3: Choose the carrier's vehicle with minimal costs. The costs are calculated as follows:

- $\sigma = \min_q\{\mu_{k_I}^q, \nu_{k_{II}}, \omega_{k_{III}}\}$.

For the minimizing carrier and vehicle, do the following:
Carrier I: add additional undelivered batches to the vehicle for the destination (if it can carry), add another node to the vehicle route (a_j^q), mark laden batches as delivered. Update the set of delivered batches Y^q for the vehicle q with laden batches.
Carrier II: add additional undelivered batches to the destination on the condition that their delivery is cheaper than if they were delivered by Carrier III, add another node to the vehicle route (b_j), mark laden batch as delivered. Update the set of delivered batches $Y^{\mathbf{s}}$ with laden batches.
Carrier III: mark laden batch as delivered. Update the set of delivered batches $Y^{\mathbf{t}}$ with laden batches.

If there is still an undelivered batch, take the first one and proceed to Step 2, otherwise the processing continues with Step 4.

Step 4: Find better loading of Carrier I vehicle r. For vehicle r of Carrier I find a new loading.

- $Y_i^r = \{k\}$, where $k \in Y^r \cup Y^{\mathbf{s}} \cup Y^{\mathbf{t}} \wedge \sum_k w_k \leq c_r \wedge i$ is index of calculated set of new possible loading (number of indices is reduced by algorithm).
- $W_i^r = \{\sum_{k, k \in Y_i^r} w_k\}$.
- $Y_{new}^r = Y_i^r$, where $W_i^r = \max\{W_j^r\}$, where j is index of calculated set of new possible loading.

If the new loading is not higher, take next vehicle and proceed to Step 4, otherwise the processing continues with Step 5.

Step 5: Calculate new route of Carrier I vehicle r. For vehicle r of Carrier I a new route is calculated by using insert heuristic and adding undelivered batches to the vehicle for the served node.
The costs for vehicle and each undelivered batches from Y_{new}^r are calculated as follows:

- $\mu_k^r = \min_{a_j}\{p_r(d_{a_j,i_k} + d_{i_k,a_{j+1}} - d_{a_j,a_{j+1}})\}$ where $a_j \in (a_j^r) \wedge j < N_r$, $i_k =$ the destination node for batch k,

For the minimizing carrier and vehicle, do the following:
Carrier I: add additional undelivered batches to the vehicle for the destination (if it can carry), add another node to the vehicle route (a_j^r), mark laden batches as delivered. Update the set of delivered batches Y^r for the vehicle r with laden batches.
Repeat this steps until all batches from Y_{new}^r are delivered.
If new route is not cheaper than previous, take next vehicle and proceed to Step 4, otherwise the processing continues with Step 6.

Step 6: Calculate new route of Carrier II and Carrier III. Undelivered batches from $Y^r \cup Y^{\mathbf{s}} \cup Y^{\mathbf{t}}$ are assigned to Carrier II and Carrier III by using the methods from Step 2 and Step 3.
The costs for each vehicle and each undelivered batche from $Y^r \cup Y^{\mathbf{s}} \cup Y^{\mathbf{t}}$ are calculated as follows:

- $\nu_k = \min_{b_j}\{p_{\text{Carr\_II}}(d_{b_j,i_k} + d_{i_k,b_{j+1}} - d_{b_j,b_{j+1}}) + w_k r_{\text{Carr\_II}}\}$ where $b_j \in (b_j) \wedge j < M$, $i_k =$ the destination node for batch k,
- $\omega_k = w_k r_{\text{Carr\_III}}$.

For each vehicle minimal costs are selected over all undelivered batches:

- $\nu_{k_{II}} = \min_k\{\nu_k\}$,
- $\omega_{k_{III}} = \min_k\{\omega_k\}$.

The costs are calculated as follows:

- $\sigma = \min\{\nu_{k_{II}}, \omega_{k_{III}}\}$.

For the minimizing carrier and vehicle, do the following:
Carrier II: add additional undelivered batches to the destination on the condition that their delivery is cheaper than if they were delivered by Carrier III, add another node to the vehicle route (b_j), mark laden batch as delivered. Update the set of delivered batches $Y^{\mathbf{s}}$ with laden batches.
Carrier III: mark laden batch as delivered. Update the set of delivered batches $Y^{\mathbf{t}}$ with laden batches.
Repeat this steps until all batches from $Y^r \cup Y^{\mathbf{s}} \cup Y^{\mathbf{t}}$ are delivered.
If there is still a vehicle that is not optimized, take the next one and proceed to Step 4, otherwise the processing will stop.

3.3 Heuristic V

Heuristics V joins the algorithms of both previous Heuristics III and Heuristic IV.

4 Performance of the heuristics

For testing of the model and heuristics, the same datasets were used as in the previous paper, expanded by Example 2 v3, in which costs for all vehicles of Carrier I were increased. The description of the parameters is given in Table 1. The model and heuristics were tested on the computer with next configuration: Intel Core i5-4200U CPU @ 1.60 GHz 2.30 GHz, 4 GB RAM, Windows 10 (x64). The ILP model is solved by CPLEX 12.6 and both heuristics are implemented in Python 3.4.

	vert's	**batches**	**No. I**	**var's**	**constr's**
Example 1	5	10	2	115	96
Example 1 v2	5	10	2	115	96
Example 1 v3	5	10	2	115	96
Example 2	10	20	5	740	693
Example 2 v2	10	20	5	740	693
Example 2 v3	10	20	5	740	693
Example 3	30	40	8	8 500	8 446

Table 1 ***Summary data for test instances.*** *The columns stand for the number of nodes, number of batches, number of vehicles of Carrier I, number of variables and number of constraints of the ILP model. In Ex. 1, batches are sorted at random, in Ex. 1-v2, batches are sorted in the descending order according to the weights, in Ex. 1-v3, batches are sorted by distances to the target node. Ex. 2 v2 is a modification of Ex. 2 with lower per-km costs of Carrier III, in Ex. 2-v2, costs of Carrier III are lower than Carrier II, and in Ex. 2-v3, costs for all vehicles of Carrier I are increased.*

The processing results are shown in Table 3. For Heuristics III and Heuristic V, calculations were repeated 10000 times. The results obtained were the ones with the lowest costs. Table 2 contains computational time of the individual methods.

	Cplex	**Heur. I**	**Heur. II**	**Heur. III**	**Heur. IV**	**Heur. V**
Example 1	0.840	0.212	0.213	6.460	0.005	8.681
Example 1 v2	0.130	0.243	0.202	7.033	0.004	8.898
Example 1 v3	0.130	0.196	0.194	6.875	0.004	8.906
Example 2	520.450	0.501	0.655	67.091	0.023	87.592
Example 2 v2	315.740	0.455	0.653	71.160	0.020	87.120
Example 2 v3	940.050	0.478	0.586	68.679	0.017	79.924
Example 3	11 937.660	0.912	1.205	762.881	0.155	1 011.674

Table 2 ***Computational time:*** *the values of computational time for CPLEX and Heuristics (in seconds).*

	Cplex		**Heur. I**	**Heur. II**	**Heur. III**	**Heur. IV**	**Heur. V**
Example 1	68 100	costs	71 343	70 283	70 283	68 100	68 100
		H/C	1.048	1.032	1.032	1	1
Example 1 v2	68 100	costs	74 833	70 283	70 283	68 100	68 100
		H/C	1.099	1.032	1.032	1	1
Example 1 v3	68 100	costs	73 221	70 283	70 283	68 100	68 100
		H/C	1.075	1.032	1.032	1	1
Example 2	128 187	costs	162 366	133 790	131 200	132 796	128 812
		H/C	1.267	1.044	1.024	1.036	1.005
Example 2 v2	67 988	costs	78 581	70 928	69 539	70 125	68 961
		H/C	1.156	1.043	1.029	1.031	1.014
Example 2 v3	139 578	costs	164 003	146 501	142 519	145 513	140 690
		H/C	1.175	1.050	1.021	1.043	1.008
Example 3	≈334 237	costs	360 311	346 349	344 209	344 666	339 532
		H/C	1.078	1.036	1.030	1.031	1.016

Table 3 ***Results of experiments:*** *the optimal values of objective function found by CPLEX (in Ex. 3, only a lower bound was available due to excessive memory requirements), the values (costs) found by Heuristics and the corresponding ratios (H/C).*

5 Conclusions

We consider a version of Vehicle Routing with three types of carriers using different forms of cost functions as was built in previous working paper [6]. We defined three new heuristics, where Heuristics III modifies Heuristic II by adding randomized step, Heuristics IV modifies Heuristic II by adding algorithm for finding better solution of vehicle loading of Carrier I vehicles and Heuristics V joins the algorithms of previous Heuristics III and Heuristic IV. The newly created heuristics show better results compared to Heuristics I and II. The best one is combined Heuristic V which generates the optimal solution for smaller examples and constructs feasible solution just about 0.5% – 1.6% worse compared to the true optimal value in other cases.

Acknowledgements

The work of M.C. and M.R. was supported by the Czech Science Foundation under Grant P16-00408S.

References

[1] Braekers, K., Ramaekers, K. & Van Nieuwenhuyse, I. (2016). The vehicle routing problem: State of the art classification and review. *Computers & Industrial Engineering*, 99, 300–313.

[2] Campbell, A.M. & Savelsbergh, M. (2004). Efficient insertion heuristics for vehicle routing and scheduling problems. *Transportation Science*, 38 (3), 369–378.

[3] Chu, Ch-W. (2005). A heuristic algorithm for the truckload and less-than-truckload problem. *European Journal of Operations Research*, 165, 657–667.

[4] Eksioglu, B., Vural, A. V. & Reisman A. (2009). The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering*, 57, 1472–1483.

[5] Lin, S. & Kernighan, B. W. (1973). An Effective Heuristic Algorithm for the Traveling Salesman Problem. *Operations Research*, 21, 498–516.

[6] Štourač, P., Černý, M. & Rada, M. (2018). Two Heuristics for Vehicle Routing with Three Kinds of Carriers: A Computational Study. *Submitted.*

Ordering of fuzzy quantities with respect to a fuzzy benchmark – how the shape of the fuzzy benchmark and the choice of distance/similarity affect the ordering

Tomáš Talášek[1], Jan Stoklasa[2]

Abstract. To order several outputs of a model represented by fuzzy numbers, we can define a reference outcome of the model called benchmark (e.g. a fuzzy singleton when ideals are used as benchmarks). Then the distance (or similarity) between this reference outcome and each of the fuzzy outputs is used for the ordering of the outputs of a model. In many cases, however, the benchmark is represented by a fuzzy number i.e. when an expert estimate of the benchmark is given, or when predictions of future values are considered.
This paper investigates the consequences of using fuzzy benchmarks for the ordering of fuzzy numbers. The paper studies if and how the use of a fuzzy benchmark with different cardinality may affect the final ordering of fuzzy numbers with respect to a chosen distance/similarity of fuzzy numbers. Different sets of fuzzy numbers representing different outputs of models (e.g. fuzzy net present values) are ordered by several distances/similarities of fuzzy numbers while the definition of the fuzzy benchmark changes cardinality-wise. Based on the analysis of the results and their graphical summaries we identify distances/similarities suitable for use with fuzzy benchmarks.

Keywords: Ordering, fuzzy numbers, fuzzy benchmark, similarity, distance.

JEL classification: D81, C44
AMS classification: 90B50, 91B06

1 Introduction

In fuzzy multiple-criteria evaluation it might be difficult to select the best alternative, since the evaluations of the alternatives are frequently represented by fuzzy numbers (or even collections of fuzzy numbers). A comparison with an "ideal" value (see e.g. [2]) could be used to obtain the ordering of the alternatives. Yet for this the "ideal" has to be defined and an appropriate measure of the distance of the fuzzy-valued evaluation from (or its similarity with) the "ideal evaluation" has to be chosen. Even though a non-fuzzy ideal value seems to be a natural choice, many authors decide to use fuzzy ideals (i.e. ideals represented by a fuzzy number) in various mathematical methods including TOPSIS [5] or selection of human resources [7]. The reasons for this choice vary from the inability of some distance/similarity measures of fuzzy numbers to work with fuzzy singletons to the simple statement that since all the evaluations are fuzzy, the ideal should be fuzzy as well. We are not analyzing the reasonability of this choice in the present paper, though. Our aim is to investigate what are the possible consequences of the use of fuzzy ideal in the process of determination of the ordering of alternatives represented by fuzzy evaluations. We depart from the widely accepted alpha-cut ordering of fuzzy numbers and show that under some circumstances even this "natural" ordering of fuzzy numbers can be contradicted by the results of a ordering procedure based on the similarity of fuzzy numbers and using a fuzzy ideal. Even more importantly we show that under the same similarity of fuzzy numbers the ordering of the given fuzzy evaluations can change simply as a result of the change of the uncertainty of the fuzzy ideal.

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 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.

A fuzzy number is a fuzzy set A on the set of real numbers which satisfies the following conditions: (1)

[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.

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) Supp(A) is bounded. A fuzzy number A is said to be defined on $[a,b]$, if 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 $[a_1,a_4] =$ 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)) $= [\underline{a}(0), \overline{a}(0)]$. Given two fuzzy numbers A and B on the same universe U, their natural ordering can be defined based on their α-cuts in the following way: if $A_\alpha \leq B_\alpha$ for all $\alpha \in (0,1]$, then A is less or equal to B, denoted $A \leq_\alpha B$.

The 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 $a_2 = a_3$.

Let A be a fuzzy number on $[a,b]$. Then the *cardinality* of A, denoted by Card(A) is computed by Card(A) $= \int_a^b A(x)dx$. If $a_1 \neq a_4$, then the *center of gravity* of A denoted by COG(A) is computed by COG(A) $= \int_a^b xA(x)dx/$Card(A). More details on fuzzy numbers and computations with them can be found for example in [3].

3 Ordering of outputs of mathematical model with respect to a benchmark

Let $O = \{O_1,\ldots,O_n\}$ be a set of n outputs of mathematical model (evaluations), where O_i, $i = 1,\ldots,n$, is a fuzzy number and let the fuzzy number B be a reference outcome of the model called benchmark (i.e. the desired output of the model, or in other words a (fuzzy) ideal value). The goal is to order the outputs O_i of the mathematical model with respect to the chosen benchmark B. To reach this goal a distance $d(B,O_i)$ (or similarity $s(B,O_i)$) between the benchmark B and each output O_i must be computed. Then, the outputs of the model can be ordered with respect to the computed distances in the ascending order (in the case of similarity, the outputs are ordered in the descending order).

It is important to keep in mind that the selected distance/similarity of fuzzy numbers affects the output of the final ordering, because different distances/similarities focus on different attributes of the fuzzy numbers to be ordered. The definition of the benchmark also plays a significant role (see the following sections). For the purpose of this paper we investigate two distances of fuzzy numbers (d_1 and d_2) and two similarities of fuzzy numbers (s_1 and s_2) - see their definitions for the case of fuzzy numbers C and D defined on the same universe (analogous measures were analyzed e.g. in [8, 10]):

- A *modified Bhattacharyya distance* [1]

$$d_1(C,D) = \left[1 - \int_U (C^*(x)\cdot D^*(x))^{1/2}dx\right]^{1/2}, \tag{1}$$

  where $C^*(x) = C(x)/\text{Card}(C(x))$ and $D^*(x) = D(x)/\text{Card}(D(x))$. Note that this distance measure requires Card($C(x)$) $\neq 0$ and Card($D(x)$) $\neq 0$, i.e. neither the fuzzy evaluation nor the fuzzy ideal can be represented by a fuzzy singleton.
- A *dissemblance index* [6]

$$d_2(C,D) = \int_0^1 |\underline{c}(\alpha) - \underline{d}(\alpha)| + |\overline{c}(\alpha) - \overline{d}(\alpha)|\, d\alpha, \tag{2}$$

- A Weis and Chens *similarity measure* [11]

$$s_1(C,D) = \left(1 - \frac{\sum_{i=1}^4 |c_i - d_i|}{4}\right)\cdot \frac{\min\{Pe(C),Pe(D)\} + \min\{\text{hgt}(C),\text{hgt}(D)\}}{\max\{Pe(C),Pe(D)\} + \max\{\text{hgt}(C),\text{hgt}(D)\}}, \tag{3}$$

  where $Pe(C) = \sqrt{(c_1-c_2)^2 + (\text{hgt}(C))^2} + \sqrt{(c_3-c_4)^2 + (\text{hgt}(C))^2} + (c_3-c_2) + (c_4-c_1)$, $Pe(D)$ is defined analogically,
- A Hejazi and Doostparast *similarity measure* [4]

$$s_2(C,D) = \left(1 - \frac{\sum_{i=1}^4 |c_i - d_i|}{4}\right)\cdot \frac{\min\{Pe(C),Pe(D)\}}{\max\{Pe(C),Pe(D)\}}\cdot \frac{\min\{Ar(C),Ar(D)\} + \min\{\text{hgt}(C),\text{hgt}(D)\}}{\max\{Ar(C),Ar(D)\} + \max\{\text{hgt}(C),\text{hgt}(D)\}}, \tag{4}$$

  where $Ar(C) = \frac{1}{2}\text{hgt}(C)(c_3 - c_2 + c_4 - c_1)$, $Ar(D)$ is defined analogically and $Pe(C)$ and $Pe(D)$ are computed identically as in the previous method.

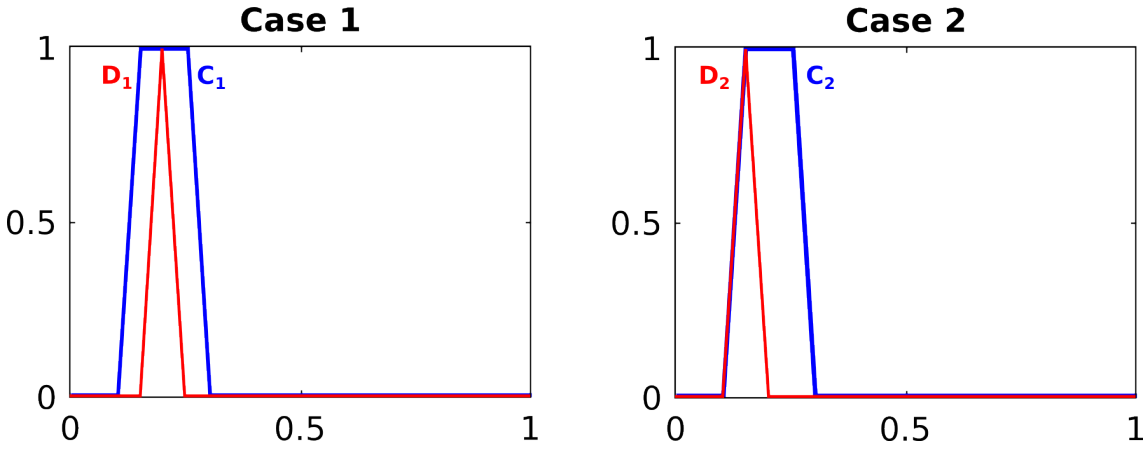


Figure 1 The selected pairs of fuzzy evaluations to be ordered with respect to a chosen fuzzy ideal. Left subfigure represents a case where the two fuzzy evaluations cannot be ordered based on the natural α-cut ordering, Right subfigure represents a case where the $D_2 \leq_\alpha C_2$. In both cases the shape of the red fuzzy number is the same (it is a symmetrical triangular fuzzy number with the same cardinality), just the location (i.e. center of gravity) differs. The blue fuzzy numbers are identical in both subfigures.

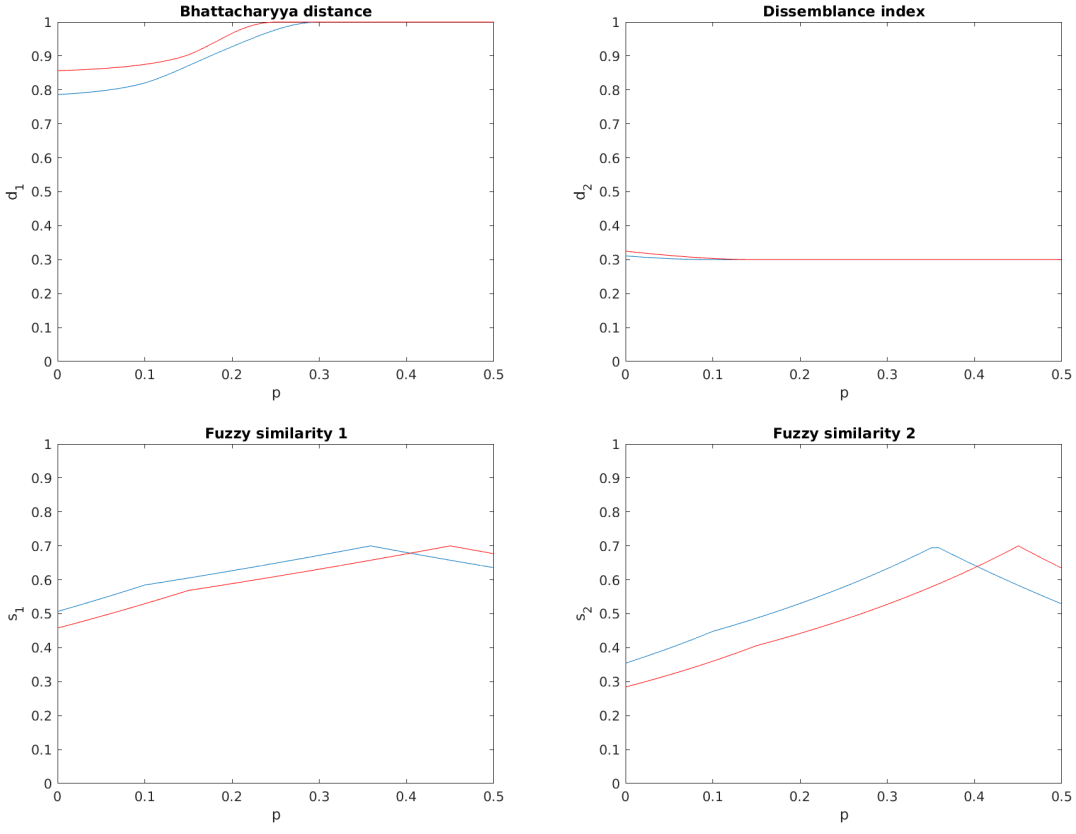


Figure 2 Results of the analysis for C_1 (blue curves) and D_1 (red curves). Top left subfigure summarizes $d_1(C_1, B_p)$ in blue and $d_1(D_1, B_p)$ in red. Top right subfigure summarizes $d_2(C_1, B_p)$ in blue and $d_2(D_1, B_p)$ in red. In both cases whichever curve is the lowest represents the preferred alternative/evaluation w.r.t. the given value of p. Bottom left subfigure summarizes $s_1(C_1, B_p)$ in blue and $s_1(D_1, B_p)$ in red. Bottom right subfigure summarizes $s_2(C_1, B_p)$ in blue and $s_2(D_1, B_p)$ in red. In both cases whichever curve is the highest represents the preferred alternative/evaluation w.r.t. the given value of p.

Let us, for the sake of simplicity, consider just two pairs of fuzzy evaluations to be ordered. The selected two pairs are depicted in Figure 1 (similar pairs of fuzzy numbers are analyzed e.g. in [11]). In the next section we apply the four above-mentioned distance/similarity measures of fuzzy numbers to obtain the ordering of C_1 and D_1 and C_2 and D_2 respectively.

As a benchmark we choose to use a parameterized symmetrical triangular fuzzy number $B_p = (p, 0.5, 0.5, 1 - p), p \in [0, 0.5)$ with a center of gravity larger than any of the significant values of the fuzzy numbers to be ordered; without any loss of generality we set $\text{COG}(B_p) = 0.5$. The cardinality of this fuzzy benchmark is dependent on the choice of the value of the parameter p (another analysis suggesting to use parametrized fuzzy number to explore the behaviour of fuzzy distance can be found in [9]). In the next chapter we analyze what ordering is suggested by (1) - (4) for C_1 and D_1 and C_2 and D_2 for any choice of $p \in [0, 0.5)$.

4 Analysis of the performance of the selected distance/similarity measures of fuzzy numbers

Two pairs of fuzzy numbers representing the expected outputs of mathematical model were chosen (see Figure 1): $C_1 = (0.10, 0.15, 0.25, 0.3), D_1 = (0.15, 0.20, 0.20, 0.25)$ in its left subfigure and $C_2 = (0.10, 0.15, 0.25, 0.3)$, $D_2 = (0.10, 0.15, 0.15, 0.20)$ in its right subfigure. Please note, that fuzzy numbers C_1 and C_2 represent the same trapezoidal fuzzy number and fuzzy numbers D_1 and D_2 have same shape but different center of gravity. Also note, that fuzzy numbers from the first pair have identical centers of gravity. The left subfigure of Figure 1 thus represents a case where the two fuzzy evaluations cannot be ordered based on their α-cuts, while in the right subfigure we clearly have $D_2 \leq_\alpha C_2$. We would thus expect that all four chosen distance/similarity measures will thus provide the same ordering for C_2 and D_2, even regardless of the definition of the benchmark B_p. We have calculated the distances/similarities of the fuzzy evaluations to all the values of the fuzzy ideal $B_p, p \in [0, 0.5)$ for both cases from Figure 1 and present the results graphically in Figures 2 and 3. Note, that for the d_1 and d_2 distances (top 2 subfigures in both figures) the lower curve represents the preferred evaluation, while for the similarities s_1 and s_2 the higher curve represents the preferred evaluation for the given value of p.

In the Case 1 (fuzzy evaluations C_1 and D_1) both distances of fuzzy numbers favor the trapezoidal fuzzy number C_1 before triangular fuzzy number D_1. However, this applies only for the cases, where the benchmark B_p has higher cardinality (p is close to 0). When the value of p crosses a certain threshold (different for both distances), both distances stop discriminating between C_1 and D_1. In other words when the cardinality of B_p is low, which results in an empty intersection of B_p with C_1 and D_1, both fuzzy evaluations are considered equally distant from the ideal. This is especially evident in the case of the dissemblace index (d_2). We, however, need to keep in mind that C_1 and D_1 have the same center of gravity and are the very prototype of a pair of fuzzy numbers that cannot be ordered based on the α-cut ordering. On the other hand similarity s_1 and s_2 distinguish between the fuzzy evaluations C_1 and D_1 even in the cases when p is close to 0.5. However, in both cases there is a value of p (around 0.4), when the ordering of C_1 and D_1 switches. For p lower then this threshold, both methods prefer the trapezoidal fuzzy number C_1 (as well as in the case of d_1 and d_2), but when the cardinality of the benchmark becomes lower, both method start to prefer D_1 over C_1. This behaviour of the similarity measures can be attributed to their focus on the shape of the fuzzy numbers being compared - when the cardinality of the fuzzy ideal B_p gets lower, its shape gets closer to the shape of the triangular C_1 evaluation. This would suggest that the similarity measures focusing on the shape are not the best choice for the ordering of fuzzy evaluations. Their use can be, however, justified, as long as cardinality (or the amount of uncertainty) of the fuzzy ideal should be well matched by the best alternative.

For the Case 2 which is represented by the fuzzy numbers C_2 and D_2 to be ordered, the modified Bhattacharyya distance d_1 provides similar results as in the previous case – the trapezoidal fuzzy number C_2 is preferred over the triangular fuzzy number D_2. Unlike in the first case the dissemblace index now favors the trapezoidal fuzzy number C_2 for any value of p. This behaviour is expected because D_2 is a subset of fuzzy number C_2 and $\text{COG}(D_2) < \text{COG}(C_2)$. The C_2 is also preferred over D_2 if the similarity s_1 is used. And again this preference is independent on the value of p. However, for the similarity s_2 there again exists a threshold, where the preference is switched. In other words from a certain value of p the similarity s_2 gives results directly opposite to those suggested by the α-cut ordering of the fuzzy evaluations C_2 and D_2. This is very counterintuitive, but it can still be explained by the focus of s_2 on the shape of the fuzzy numbers that are being compared.

From the above text can be conclude, that the Bhattacharyya distance d_1 exhibits consistent behaviour, but can only order fuzzy numbers, if the benchmark B_p has higher cardinality (p is close to zero). In fact d_1 discriminates (returns a value different from 1) only if the intersection of the fuzzy evaluation with the fuzzy ideal is nonempty. This might not be a desired property. The Dissemblance index d_2 may overcome this disadvantage of d_1 in some cases, but it is still possible that the distances may coincide. However, as shown in Case 1, this happens in the "difficult to resolve" cases where the inability to distinguish two evaluations might actually be the correct answer we are looking for (in some cases the values simply cannot be ordered without additional information).

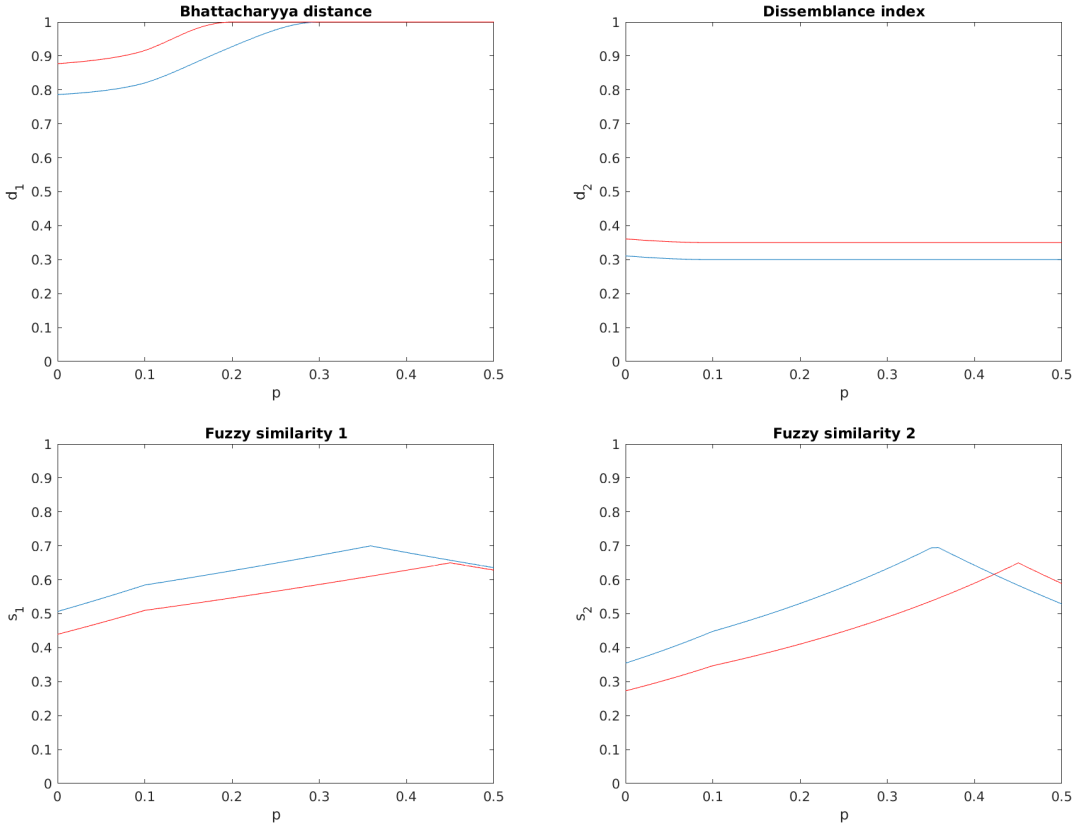


Figure 3 Results of the analysis for C_2 and D_2. Top left subfigure summarizes $d_1(C_2, B_p)$ in blue and $d_1(D_2, B_p)$ in red. Top right subfigure summarizes $d_2(C_2, B_p)$ in blue and $d_2(D_2, B_p)$ in red. In both cases whichever curve is the lowest represents the preferred alternative/evaluation w.r.t. the given value of p. Bottom left subfigure summarizes $s_1(C_2, B_p)$ in blue and $s_1(D_2, B_p)$ in red. Bottom right subfigure summarizes $s_2(C_2, B_p)$ in blue and $s_2(D_2, B_p)$ in red. In both cases whichever curve is the highest represents the preferred alternative/evaluation w.r.t. the given value of p.

If the similarity of fuzzy numbers is used, it is important to take into consideration, that for different cardinality of benchmark B_p the ordering could be different; it can even be in conflict with the natural ordering of fuzzy numbers based on the α-cuts. This is caused due to the construction of these similarities, because they take into consideration the shape of fuzzy numbers. This is especially visible in the case of s_2.

Overall the chosen distance/similarity measures seem to perform better for low values of p, i.e. for fuzzy ideals with high cardinality. The less uncertain the fuzzy ideal is, the more problems seem to appear - ranging from the loss of discrimination power in d_1 to a complete switch in the ordering of the alternatives when s_2 is used. This is an interesting finding - mainly because fuzzy ideals are often derived by fuzzification from crisp (non-fuzzy) ideal values just to allow for the use of distances/similarities of fuzzy numbers. Our findings suggest that just adding a little uncertainty can be counterproductive as long as the ordering of alternatives based on their fuzzy evaluations is needed. Out of the four analyzed distance/similarity measured the dissemblance index (d_2) seems to be performing best and seems to be the least dependent on the definition of the fuzzy ideal.

5 Conclusion

In the paper we investigate the reasonability of using fuzzy ideals in connection with two chosen distances of fuzzy numbers and two chosen similarities of fuzzy numbers for the purpose of ordering of fuzzy evaluations. We show that in many circumstances the resulting ordering depends not only on the choice of the distance/similarity measure, but also on the definition of the fuzzy ideal. Surprisingly, low-uncertain fuzzy ideals do not seem to provide appropriate results. The only distance/similarity measure that did not exhibit significant drawbacks under any definition of the fuzzy ideal is the dissemblance index. Even though this study focuses on just four chosen distance/similarity measures of fuzzy numbers, it still provides valuable insights into the use of fuzzy ideals in the evaluation setting.

Acknowledgements

Supported by the grant No. IGA_FF_2018_002 of the Internal Grant Agency of Palacký University Olomouc.

References

[1] Bonissone, P. P.: A pattern recognition approach to the problem of linguistic approximation in system analysis. *Proceedings of the IEEE International Conference on Cybernetics and Society* (1979), 793–798.

[2] Deng, H.: Comparing and ranking fuzzy numbers using ideal solutions. *Applied Mathematical Modelling* **38** (2014), 1638–1646.

[3] Dubois, D., and Prade, H.: *Fuzzy sets and systems: theory and applications*. Nf. Academic Press, New York, 1980.

[4] Hejazi, S., Doostparast, A., and Hosseini, S.: An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers. *Expert Systems with Applications* **38** (2011), 9179–9185.

[5] Javadian, N., Kazemi, M., Khaksar-Haghani, F., Amiri-Aref, M., and Kia, R.: A general fuzzy TOPSIS based on new fuzzy positive and negative ideal solution. In: *2009 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, Hong Kong, China, 2271–2274.

[6] Kaufmann, A., and Gupta, M. M.: *Introduction to Fuzzy Arithmetic: Theory and Applications*. Electrical-Computer Science and Engineering Series. Van Nostrand Reinhold, New York, 1985.

[7] Luukka, P., and Collan, M.: Fuzzy Scorecards, FHOWA, and a New Fuzzy Similarity Based Ranking Method for Selection of Human Resources. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 601–606.

[8] 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* **29** (2017), 485–503.

[9] Talášek, T., and Stoklasa, J.: Selection of tools for managerial decision support the identification of methods of choice in linguistic approximation. In: *Proceedings of the international scientific conference Knowledge for Market Use 2017* (Slavíčková, P., ed.). Palacký University, Olomouc, Olomouc, 606–613.

[10] 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* (Kocourek, A., and Vavroušek, M., eds.). Technická Univerzita v Liberci, Liberec, 845–850.

[11] Wei, S.-H., and Chen, S.-M.: A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *Expert Systems with Applications* **36** (2009), 589–598.

Facility Location Problems with Semi-fixed Costs and Time Availability

Dušan Teichmann[1], Alexandra Dvořáčková[2], Denisa Mocková[3]

Abstract. Logistic chains must be planned efficiently and economically. Effectiveness demands goods being delivered to the required quality while economic criteria ask for delivery of goods at minimal cost.
The following article considers the economic facets of the delivery process as expressed against the total costs of the process. Operation research methods, specifically mathematical programming models, will be used to calculate economic optimization scenarios. A range of distribution problems can be calculated by means of mathematical programming. Mathematical models applied to various problems may differ: some variants of the problems include locations of warehouses which are known; for others warehouse locations must be determined. In addition, particular customers may require deliveries from different locations at the same time; other customers may restrict deliveries to a single supplier.
This article examines a problem in which warehouse locations must be determined (potential locations are known) and each customer may only take delivery from a single (operational) storage point. This type of problems is termed a facility location problem. In the course of this problem it shall be assumed that storage points at each location may be set up in any of multiple variants, which differ in capacity and subsequent associated setup costs. The optimization criteria taken into account will be the total costs for setting up a warehouse, its maintenance and cargo delivering.
In addition to these cost criteria, the problem will take into account a second variable expressing the worst (maximum) time availability. A requirement for minimization will be set for both criteria. The STEM methodology will be used for calculations.

Keywords: Facility Location Problem, Linear Programming, Operation Research.

JEL Classification: C44
AMS Classification: 90C11

1 Introduction: Motivation for addressing the problem

Location problems play an important role in distribution logistics. Their objective is to optimize positions of warehouses, which are to be placed in predefined locations in a given territory and to create attraction zones for them. The most frequently selected optimization criterion is the total running costs of the system. In principle, standard models of location problems assume that the fixed costs for individual locations do not change. However, this assumption may be inadequate in real-life logistics practice. It is mainly because there can be more variants available when deciding about the location of a warehouse. These variants are distinguished by their capacities and the correspondent various values of the fixed costs. The changing fixed costs are called semi-fixed costs in economic literature; see for example [5]. Apart from the total costs, other criteria affecting the quality of logistic services should be taken into consideration [2], [4], [6] or [7]. These other criteria include, for instance, the time availability of the supplied goods. Frequently, customers do not operate their own warehouse and they expect the time from the moment of ordering to the moment of delivery to be as short as possible. The level of the time availability will be modelled and evaluated by means of the value of the worst (maximum) time availability.

The presented article contains a proposal of a mathematical model for the location problem that incorporates the semi-fixed costs for the operation of warehouses. In addition, it takes into consideration the time availability of the customers which is needed for the transfer of the good from warehouses. The model includes two objective functions. The first objective function represents the total cost of supplying the customers while the

[1] VŠB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu, 15/2172, Ostrava-Poruba, dusan.teichmann@vsb.cz.
[2] Czech Technical University in Prague, Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Konviktská, 20, Praha 1, dvoraale@fd.cvut.cz.
[3] Czech Technical University in Prague, Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Konviktská, 20, Praha 1, mockova@fd.cvut.cz.

second function represents the value of the worst time availability. The aim of the optimization process is to minimize values of both objective functions while respecting the constraints.

2 Mathematical formulation of problem and mathematical model

This chapter formulates the assignment for the mathematical solution of the problem and proposes a mathematical model. In the transport network there is n number of customers and m number of locations in which the warehouses can be located. The maximum capacity a_i for a certain planning period for each warehouse $i = 1,...,m$ is known (provided the warehouse is to be operated there). For each location $i = 1,...,m$ a set of intervals K_i is defined; this set represents the volume of the goods distributed through this location, which results in various operational costs for the warehouse. The previously mentioned intervals are assumed to be disjunctive and continuous. I.e. $\langle 0; h_{i1} \rangle \cup (h_{i1}; h_{i2} \rangle \cup ... \cup (h_{ip_i-1}; h_{ip_i} \rangle$ and then let $h_{p_i} = a_i$. Semi-fixed costs f_{ik} for each location $i = 1,...,m$ and interval $k = 1,..., p_i$ are known. Let us assume that $f_{ik-1} < f_{ik}$ for $i = 1,...,m$ and $k = 2,..., p_i$ (this assumption indicates that the fixed costs will increase with the increasing volume of goods distributed via the warehouse). For each customer $j = 1,...,n$ the requirement b_j is defined for the same planning period. In addition, the unit costs of supply to the customer $j = 1,...,n$ from the warehouse in the location $i = 1,...,m$ are known; these are referred to as c_{ij} further in the text. Let us assume that the customer cannot be supplied from several warehouses at the same time. In addition, let us assume that each customer can be supplied from any operational warehouse. There is also a time availability matrix in which element t_{ij} represents the value of the customer's time availability $j = 1,...,n$ from the warehouse $i = 1,...,m$. The task is to determine a way of clearly assigning the customers to the operational warehouses in order to minimize the amount of total costs of supplying customers from the warehouses and at the same time to minimize the value of the worst (maximum) time availability.

Three groups of variables are entered into the problem. The first group includes variables that model decisions on operation of warehouses in individual locations, which will generate fixed costs for the operation of the warehouse in the location. Let us designate a variable modelling a decision on operating a warehouse in location $i = 1,...,m$ as y_{ik}, where $k = 1,..., p_i$. This variable will be bivalent. If $y_{ik} = 1$ after completing the optimization calculation, then the warehouse in the location $i = 1,...,m$ will be operated and the volume of goods distributed through this warehouse will be from the interval $k = 1,..., p_i$. If $y_{ik} = 0$ after completing the optimization calculation, then the warehouse in the location $i = 1,...,m$ with a capacity within $k = 1,..., p_i$ will not be operated. The second group of variables includes variables modelling the decision of assigning customers to attraction zones of the operated warehouses (based on their values it will be possible to identify the attraction zones). Let us designate a variable modelling the said decision by the symbol x_{ij}. This variable will also be bivalent. If $x_{ij} = 1$ after completing the optimization calculation, then the customer $j = 1,...,n$ will be supplied from the operated warehouse in the location $i = 1,...,m$. If $x_{ij} = 0$ after completing the optimization calculation, then the customer $j = 1,...,n$ will not be supplied through the location $i = 1,...,m$. The third group of variables will contain only one variable. This variable will set the upper limit for the worst time availability and it will be designated by a symbol u. The nature of this variable implies that its domain will be the set R_0^+. The mathematical model for the formulated problem is as follows:

$$\min f(x, y, u) = \sum_{i=1}^{m} \sum_{k=1}^{p_i} f_{ik} y_{ik} + \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} b_j x_{ij} \tag{1}$$

$$\min f(x, y, u) = u \tag{2}$$

subject to:

$$\sum_{i=1}^{m} x_{ij} = 1 \qquad for\ j = 1,...,n \tag{3}$$

$$\sum_{k=2}^{p_i}\left(h_{ik-1}+e\right)y_{ik-1} \le \sum_{j=1}^{n} b_j x_{ij} \qquad for\ i=1,...,m \tag{4}$$

$$\sum_{j=1}^{n} b_j x_{ij} \le \sum_{k=1}^{p_i} h_{ik} y_{ik} \qquad for\ i=1,...,m \tag{5}$$

$$\sum_{k=1}^{p_i} y_{ik} \le 1 \qquad for\ i=1,...,m \tag{6}$$

$$t_{ij} x_{ij} \le u \qquad for\ i=1,...,m\ and\ j=1,...,n \tag{7}$$

$$x_{ij} \in \{0;1\} \qquad for\ i=1,...,m\ and\ j=1,...,n \tag{8}$$

$$y_{ik} \in \{0;1\} \qquad for\ i=1,...,m\ and\ k=1,...,p_i \tag{9}$$

$$u \in R_0^+ \tag{10}$$

Function (1) represents the first objective function – the total costs of the supply process. Function (2) represents the second objective function – the worst value of the time availability. The group of constraints (3) ensures that each customer is supplied. The number of constraints corresponds to the number of customers. Groups of constraints (4) and (5) ensure that the amounts of goods distributed through individual warehouses will correspond to the amount of fixed costs of operating the warehouses. In addition, the groups of constraints (4) and (5) provide logical links among variables from various groups. The groups of constraints (4) and (5) are created for every location. Therefore, their number corresponds to the number of locations. Value e in the group of constraints (4) ensures that there will be a distance between the lower bound of one interval from the upper bound of the previous interval. Group of constrains (5) further ensures that the storage capacities will be accepted. Group of constraints (6) will ensure that a maximum of one fixed cost rate will be selected for each location suitable for the operation of the warehouse. The number of constraints in this group corresponds to the number of locations. Group of constraints (7) creates links between time availability and objective function (2). The number of constraints in this group corresponds to the product of the numbers of warehouses and customers. Groups of constraints (8), (9) and (10) define the domains of values of individual variables in the model.

3 Mathematical approach to solving the proposed model

For solving the actual mathematical model the STEM method will be used. The theoretical description of the STEM method is given in publications [1] or [3]. This method is to be supplemented by the so-called cascade approach used in optimization problems of the Minimax type. The cascade approach is based on the gradual relaxation of constraints that prevent improving the value of the objective function. In models (1) – (10), the constraints of type (7) will gradually be fixed. Simultaneously, with the relaxation of constraints (7) it is necessary to modify the capacity constraints related to the warehouses (4) and (5) through which the goods are distributed in fixed flows. Operations will be made by gradual increments of the expressions by values of fixed flows on the left side for constraints of the type (4), and by the same values on the left sides of the corresponding constraints of the type (5). The entire optimization process can be described by a sequence of six steps (Steps 1 through 5 represent the initial phase of the STEM method based on which a default matrix of objective functions is obtained):

1. the mathematical model (3) – (6) and (8) – (9) with the objective function (1) is solved;
2. the mathematical model (3) – (10) with the objective function (2) and the supplementary constraint (11) is solved

$$\sum_{i=1}^{3}\sum_{k=1}^{4} f_{ik} y_{ik} + \sum_{i=1}^{3}\sum_{j=1}^{10} c_{ij} b_j x_{ij} \le C \tag{11}$$

   where C is the value of the objective function (1) calculated in Step 1,
3. the mathematical model (3) – (10) with the objective function (2) is solved,

4. the cascade approach is applied and repeated until it is no longer possible to improve the value of the objective function (2),
5. the mathematical model (3)
6. further steps of the STEM are performed; the input data for these next steps of the STEM method are the results of Steps 1, 2, 3 and 5. First, the coefficient matrix of the objective functions is compiled, and then the weights of individual criteria are calculated. Consequently, the model in which the maximum weighted deviation from the ideal variant is being minimized is solved.

Owing to the fact that the problem addressed is a model problem, let us assume that the values of both objective functions, after the first computation step of the STEM method (after the first minimization of the maximum weighted deviation from the ideal variant), will meet the customer's requirements.

4 Numerical Experiment

Let us assume a model example. In the location problem there are 3 locations suitable for setting up warehouses and 10 customers who require services. Let us assume that for each location that is suitable for setting up a warehouse there are four related values of the fixed costs, i.e. $p_i = 4$ for $i = 1,...,3$. With increasing capacity of the warehouse the fixed costs also increase. These fixed costs increase by a step change when the values of the capacity $h_{i1}, h_{i2}, ..., h_{ip_i-1} = h_{i3}$ are exceeded and then $h_{ip_i} = h_{i4} = a_i$. Information related to locations is provided in Table 1. Information on unit costs of transport in individual relations and customers' requirements (values before /) and the values of the customer's time availability from warehouses (values after /) are stated in Table 2.

Warehouse in location i	h_{i1}	h_{i2}	h_{i3}	$h_{i4} = a_i$	f_{i1}	f_{i2}	f_{i3}	f_{i4}
1	110	120	130	170	1,100	1,200	1,300	1,700
2	40	50	65	70	4,000	5,000	6,500	7,000
3	30	35	40	45	3,000	3,500	4,000	4,500

Table 1 Boundary values for changes in fixed costs and warehouse capacity

Customer j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Warehouse in location i										
1	10/3	2/5	3/4	4/6	5/12	6/5	7/4	8/3	9/8	7/9
2	11/11	3/12	5/5	7/7	9/10	4/14	6/5	5/9	3/6	8/8
3	8/8	9/19	5/5	6/6	4/4	3/13	2/12	9/9	8/18	16/16
Requirement b_j	30	35	40	45	10	15	30	20	15	30

Table 2 Unit costs of transport and customer's requirements

Step 1: Results – assigning customers to warehouses is summarized in Table 3. The total cost was $C = 12960$ monetary units.

Customer j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Warehouse in location i	1	2	1	1	1	1	3	2	2	1

Table 3 Results of Step 1

Step 2: Value of the objective function – the worst time availability was 12 units of time.

Step 3: Results assigning customers to warehouses is summarized in Table 4. The value of the objective function the worst time availability was 12 units of time.

Customer j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Warehouse in location i	3	1	2	1	3	1	1	1	1	2

Table 4 Results of Step 3

Step 4: the cascade approach was applied. Results – assigning customers to warehouses is summarized in Table 5. The first line of the Table 5 is the same as Table 4 because Step 3 can be considered as the first iteration of the cascade approach. Numbers in brackets in the first column represent numbers of iterations of the cascade approach. The symbol | means that assigning the warehouse does not change because the final assigning was achieved in the previous iteration. In addition, the tables contain values of the worst time availability for unrelaxed constraints in the course of the solution.

Customer j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	u
Warehouse assigning in location i (1)	3	1	2	1	3	1	1	1	1	2	8
Warehouse assigning in location i (2)	|	1	1	1	3	1	2	1	|	|	6
Warehouse assigning in location i (3)	|	1	2	|	3	1	1	1	|	|	5
Warehouse assigning in location i (4)	|	|	|	|	3	|	1	1	|	|	4
Warehouse assigning in location i (5)	|	|	|	|	|	|	|	1	|	|	3

Table 5 Results of Step 4

Table 6 specifies the results in individual iterations that are related to cost values according to various iterations.

Iteration	**Selected cost group of fixed costs for warehouse 1**	**Selected cost group of fixed costs for warehouse 2**	**Selected cost group of fixed costs for warehouse 3**	$\sum_{i=1}^{3}\sum_{k=1}^{4} f_{ik}y_{ik}$	$\sum_{i=1}^{3}\sum_{j=1}^{10} c_{ij}b_j x_{ij}$	C
(1)	4	4	3	12,700	1,565	14,265
(2)	4	3	3	12,200	1,455	13,655
(3)	4	4	4	13,200	1,565	14,765
(4)	4	4	4	13,200	1,565	14,765
(5)	4	4	3	12,700	1,565	14,265

Table 6 Results of Step 4 – values of objective functions

Step 5: the mathematical model (3) – (6) and (8) – (9) with objective function (1) and supplementary constrains generated in the last iteration of the cascade approach from Step 4 was solved. The result of this step is shown in the field that is the intersection of the last line and the last column in Table 6.

Step 6: further steps of the STEM method are to be performed; the input data for these next steps of the STEM are the results of Steps 1, 2, 3 and 5. Based on the results of Steps 1, 2, 4 and 5, a coefficient matrix Z of the objective functions is compiled. Since there are 2 objective functions in the problem (the total cost of supply to customers and the worst time availability) the coefficient matrix of objective functions will have a dimension of 2×2. The form of the matrix can be seen in (12).

$$Z = \begin{bmatrix} 12960 & 14265 \\ 12 & 8 \end{bmatrix} \tag{12}$$

The element in position (1,1) represents the value of the objective function (1) obtained in Step 1. The element in position (1,2) represents the value of the objective function (2) obtained in Step 2. The element in position (2,1) represents the value of the objective function (1) obtained in Step 5. Finally, the element in position (2,2) represents the value of the objective function (2) obtained in Step 3. The following procedure was used to calculate the weights of individual criteria. First, the equation (13) was solved:

$$\frac{14265-12960}{12960}\frac{\alpha}{\sqrt{12960^2+14265^2}}+\frac{12-8}{8}\frac{\alpha}{\sqrt{12^2+8^2}}=1 \tag{13}$$

from where $\alpha = 28,84007$ and then, using this value, the individual weights from relations (14) and (15) were calculated:

$$w_1 = \frac{14265-12960}{12960}\frac{28.84007}{\sqrt{12960^2+14265^2}} = 0.00015 \tag{14}$$

$$w_2 = \frac{12-8}{8}\frac{28.84007}{\sqrt{12^2+8^2}} = 0.99985 \tag{15}$$

The results were rounded to 5 decimal places. When minimizing the maximum weighted deviation from the ideal variant, the model in the form (3) – (10) supplemented by constraints (16) and (17) in the following forms is solved:

$$0.00015\left(\sum_{i=1}^{3}\sum_{k=1}^{4} f_{ik}y_{ik} + \sum_{i=1}^{3}\sum_{j=1}^{10} c_{ij}b_j x_{ij} - 12960\right) \le d \tag{16}$$

$$0.99985(u-8) \le d \tag{17}$$

supplemented by objective function (18) in the form

$$\min f(x,u,d) = d \tag{18}$$

in which the variable d represents the maximum weighted deviation from the ideal variant. Constraints (16) and (17) provide logical links among variables. Since the values of time availability are entered integers, the domain R_0^+ was changed to Z_0^+ in the constraint (10), i.e. to a set of non-negative integers. Upon solving the model the results shown in Table 7 were obtained:

Customer j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Warehouse in location i	1	1	2	1	3	1	2	3	1	1

Table 7 Results of Step 6

The total cost was 13,285 monetary units and the worst value of time availability was 9 time units. In comparison with the results of single-criterion optimizations, both values of the objective functions worsened. The total cost increased by 385 monetary units and the value of the worst time availability was extended by 1 time unit. The value of the weighted deviation after minimizing was 1.51531. As previously mentioned in the introductory commentary on the problem, the authors assume that the calculated solution is acceptable for the ordering party according to both criteria. Thus, the calculation is completed.

5 Conclusion

The presented article deals with location problems including the semi-fixed costs and time availability. Linear programming was selected to solve the problem. The article contains a linear mathematical model with two criteria – the total cost of supplying customers and the worst (maximum) time availability. Both objective functions were minimized during the optimization process. The actual solution of the mathematical problem was achieved by means of the proposed algorithm which consisted of six steps. The solving algorithm combined the well-known STEM method and the so-called cascade approach, which is used for Minimax or Maximin problems. The proposed algorithm is applied to a model problem with three possible warehouses and ten customers. For each warehouse four values of the fixed costs can be used. The fact that the values of the fixed costs increase with the increasing capacity of warehouses was taken into consideration. Based on the numerical experiments the functionality of the proposed model was verified. A compromise solution was found by means of optimizations according to individual criteria; this solution worsened the values of both objective functions.

References

[1] Benayoun, R. & de Montgolfier, J, & Tergny, J. & Larichev, O. (1971). Linear programming with multiple objective function: Step method. *Mathematical Programming*, 3, 366-375.

[2] Cappanera, P. (1999): *A Survey on Obnoxious Facility Location Problems. Technical Report* TR-99-11, University of Pisa.

[3] Fiala, P. (2006). *Modely a metody rozhodování*. Praha: Oeconomica.

[4] Jánošíková, Ľ. (2009). Reduction of a Hospital Network as a Multiple Criteria Optimization Problem. *E & M Ekonomie a Management, XII* (3), 50–57.

[5] Taylor, B.D., Garrett, M., Iseki, H. Measuring cost variability in provision of transit service. *Transportation research record*, 1735, 101-112.

[6] Verter, V.: Uncapacitated and Capacitated Facility Location Problems. In: Eiselt H., Marianov V. (eds) *Foundations of Location Analysis* (pp 25-37). Springer, 2011.

[7] Zanjirani Farahani, R., SteadieSeifi, M., Asgari, N.: Multiple criteria facility location problems: A survey. *Applied Mathematical Modelling*, 34, 1689-1709.

Probability Distribution of Military Expenditure

Martin Tejkal[1], Běla Mikulášková[2], Jiří Neubauer[3]

Abstract. The contribution deals with modeling of probability distribution of military expenditure. Analysis of relationship between military expenditure and other macroeconomic variables is topical in the defense economic literature. For this purpose, various models such as vector autoregressive, error correction or panel data models are applied. The probability distribution of analyzed variables can significantly influence the performance of estimated models. The aim of this paper is to find an appropriate probability distribution for data describing military expenditure per capita in the years 1993–2016. The statistical analysis involves data of 115 countries. The authors focus mainly on a normal, lognormal, gamma and Weibull distribution. The quality of a distribution fit is based on comparison of an empirical and a theoretical cumulative distribution function and is tested by Anderson-Darling, Cramer-von Mises and Kolmogorov-Smirnov test. In addition to the type of distribution, the development (changes) of the parameters estimates is discussed and the most sensitive test statistics is suggested. The authors compared the quantiles of the estimated probability distributions as potential risk indicators.

Keywords: distribution fitting, military expenditure, risk

JEL classification: C12, E69
AMS classification: 62F99

1 Introduction

Statistical analysis of relationship between military expenditure and other macroeconomic variables is very topical in the defense economic literature. For this purpose, various models such as vector autoregressive, error correction or panel data models are applied, see for example[10] or [6]. The probability distribution of analyzed variables can significantly influence the performance of estimated models. Normality of variables is a common and frequent assumption of many widely used models and methods. There are many normality test available whose properties differ [11]. Violation of the normality assumption can appreciably affect parameter estimates, individual tests associated with the estimated model. The problem of a probability distribution is addressed by a number of authors, for example the probability distribution of income is discussed in [2], distribution fitting of a microbiological contamination data is solved in [3], or economic growth is modeled in [12].

In this contribution, we focus on distribution of military expenditure (per capita). We would like to show that the probability distribution of military spending is not normal and should be modeled by another probability distribution, for example by a log-normal distribution. From a security perspective, it is useful to focus on the extreme values. We will pay attention to the small and, above all, high values of military spending that can be a marker of potential security risk. For this reason, we calculate the necessary quantiles of probability distributions used to model the distribution of military expenditure.

We analyze data from 1993 to 2016. According to SIPRI (Stockholm International Peace Research Institute), military expenditure include all current and capital expenditure on the armed forces (including peace keeping forces), defense ministries and other government agencies engaged in defense projects, paramilitary forces when judged to be trained, equipped and available for military operations, and military space activities. The amount of military spending in different countries varies considerably. Israel, Saudi Arabia, Oman, USA, Singapore, Kuwait, Norway Bahrain and Australia belong to countries with the largest military expenditure per capita in 2016 (military spending exceeds $ 1,000). On the other end of the scale are countries like Malawi, Madagascar, Liberia, Mozambique, Sierra Leone or Ethiopia with military expenditure less than $ 5 per head. The empirical distribution of military spending in not symmetric, it is skewed. A normal distribution is not a good model for this variable, as will be stated later.

[1]University of Defence in Brno, Department of Econometrics, Kounicova 65, Brno, Czech Republic, martin.tejkal@unob.cz
[2]University of Defence in Brno, Department of Econometrics, Kounicova 65, Brno, Czech Republic, jiri.neubauer@unob.cz
[3]University of Defence in Brno, Department of Econometrics, Kounicova 65, Brno, Czech Republic, bela.mikulaskova@unob.cz

2 Data Description

As mentioned in the introduction, the data represent military expenditure of selected countries per capita in years 1993–2016. The dataset used for the study carried out in this paper comes from the SIPRI database of military expenditure, that is available online at: `https://www.sipri.org/databases/milex`. Figure 1 shows military expenditure data of selected countries.

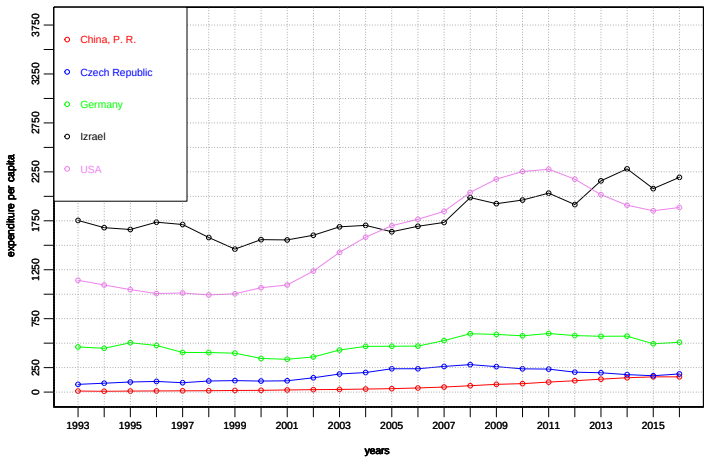


Figure 1 Military expenditure of China, the Czech Republic, Germany, Israel and USA from 1993 to 2016 (in dollars per capita)

For the purpose of the paper, from the original dataset obtained from SIPRI a selection was made. It is based on the relevance of countries included in the model and trustworthiness of the respective data. After selection the dataset contains information on 115 countries including the most prominent actors on the field of international politics like USA, China, Russian Federation, India, etc. as well as countries of former eastern block and countries of Middle East and Africa.

The time period of the data was set between the years 1993 and 2016, where the reason for the choice of the lower bound comes from the fact that data of many countries relevant for this study are not available prior to the year 1993. The upper bound corresponds to the most recent year for which the data is currently available in the SIPRI database. The goal of this paper is to study the probability distribution of military expenditure per capita in the above mentioned period by finding an appropriate distribution for military spending in each year. In case that a data entry was missing for a given country in a given year, that the country was not included into the data sample of military expenditure in that year.

Following notation will be used henceforth. Denote $n = 115$ the total number countries considered in this study and m the number of countries with missing entries in a given year j. Than $\boldsymbol{X}_j = (X_{1,j}, ..., X_{n-m,j})$ is a random sample of a size $n - m$ of military expenditure of countries in a year j.

3 Methods and Tests

It was assumed that the data sample $\boldsymbol{X}_j = (X_{1,j}, ..., X_{n-m,j})$ of military expenses of countries in a given year j had come from one of the following distributions: normal, lognormal, gamma and two-parameter Weibull distribution. For further information see [1]. The parameters μ_N, σ_N^2 of the normal distribution, μ_{LN}, σ_{LN}^2 of the lognormal distribution, the shape parameter α, the rate parameter β of the gamma distribution, the shape parameter k and the scale parameter λ of the Weibull distribution were estimated from the sample for each year j. Through the text the estimates of the parameters are denoted by hat symbol, e. g. $\hat{\mu}$ is the estimate of the parameter μ.

To obtain these estimates, maximum goodness of fit estimation as described in [8] was used, where the applied metric was Kolmogorov-Smirnov [4]. The choice of the maximum goodness of fit estimation using the Kolmogorov-Smirnov metric was done to avoid numerical computation problems that would occur while estimating parameters of the Weibull and gamma distribution when using Cramer-von Mises metric or Anderson-Darling metric in the maximum goodness of fit approach, or while using the maximum likelihood estimation approach.

Assume that the sample $\boldsymbol{X}_j = (X_{1,j}, ..., X_{n-m,j})$ comes from distribution with the distribution function $F_{\boldsymbol{X}_j}$

The following four hypotheses were tested for $j \in \{1993, 1994, \ldots, 2016\}$ and the level of significance $\alpha = 0.05$:

$$H_{01}: \quad F_{\boldsymbol{X}_j} \text{ is equal to } F_{N(\hat{\mu}_{N,j}, \hat{\sigma}^2_{N,j})}, \tag{1}$$

$$H_{02}: \quad F_{\boldsymbol{X}_j} \text{ is equal to } F_{LN(\hat{\mu}_{LN,j}, \hat{\sigma}^2_{LN,j})}, \tag{2}$$

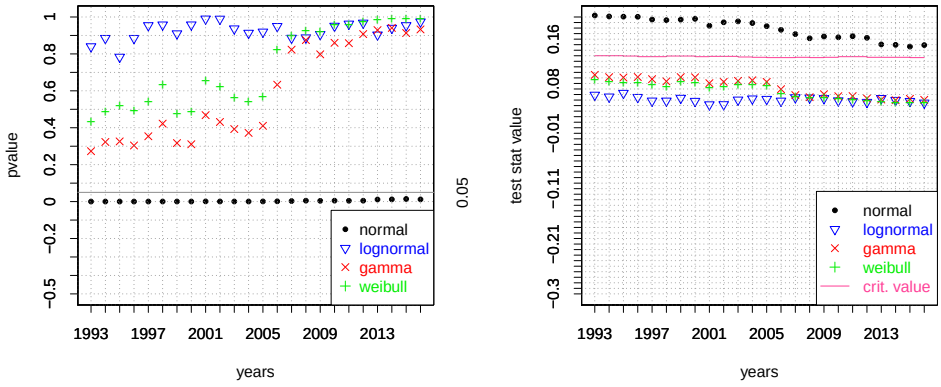
$$H_{03}: \quad F_{\boldsymbol{X}_j} \text{ is equal to } F_{G(\hat{\alpha}_j, \hat{\beta}_j)}, \tag{3}$$

$$H_{04}: \quad F_{\boldsymbol{X}_j} \text{ is equal to } F_{W(\hat{k}_j, \hat{\lambda}_j)}, \tag{4}$$

where $F_{N(\hat{\mu}_{N,j}, \hat{\sigma}^2_{N,j})}$, $F_{LN(\hat{\mu}_{LN,j}, \hat{\sigma}^2_{LN,j})}$, $F_{G(\hat{\alpha}_j, \hat{\beta}_j)}$ and $F_{W(\hat{k}_j, \hat{\lambda}_j)}$ denote in sequence the cumulative distribution functions of normal, lognormal, gamma and Weibull distributions, with the respective parameter estimates taken as the values of the actual parameters. The hypothesis testing was performed via tests based on comparing the empirical cumulative distribution function with the theoretical cumulative distribution function of the corresponding distributions using Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling metrics [4].

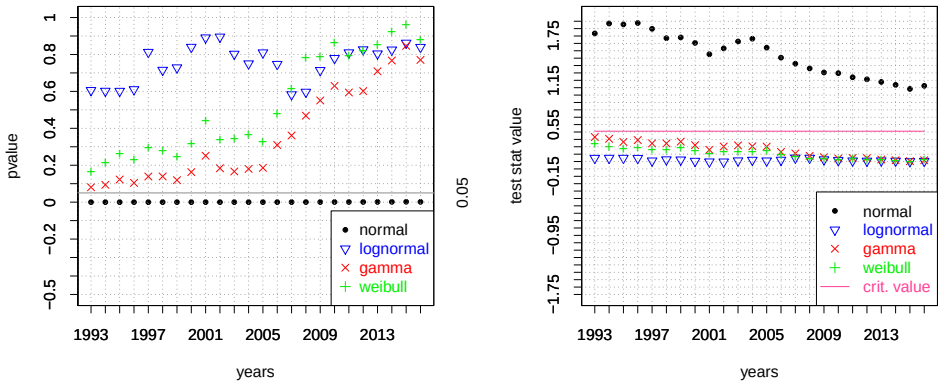
4 Results

In Figures 2a, 3a and 4a one can see how the p-values of the Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling test change for the years varying from 1993 to 2016. The gray horizontal line denotes the level of significance $\alpha = 0.05$. Figures 2b, 3b and 4b show how the test statistic values of the Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling test change in years 1993–2016. The violet-red step function displays the critical values of the test statistic for each year [5, 7, 9].



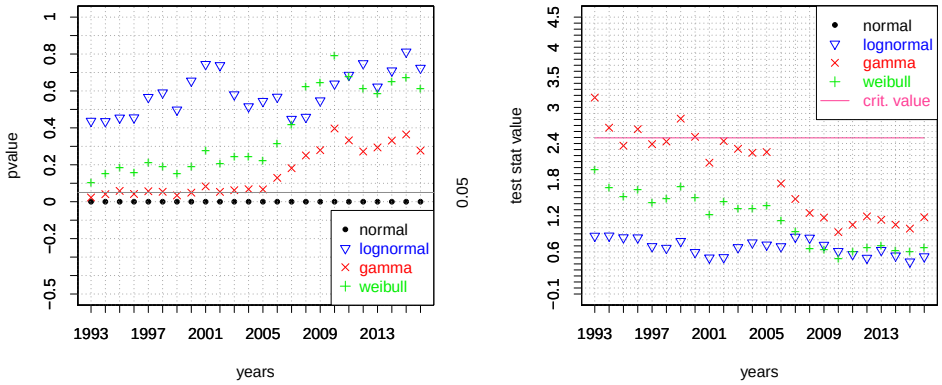
(a) P-values of Kolmogorov-Smirnov test **(b)** Values of the Kolmogorov-Smirnov test stat.

Figure 2 P-values and values of the Kolmogorov-Smirnov test statistic performed for years 1993–2016



(a) P-values of Cramer-von Mises test **(b)** Values of the Cramer-von Mises test statistic

Figure 3 P-values and values of the Cramer-von Mises test statistic performed for years 1993–2016



(a) P-values of Anderson-Darling test **(b)** Values of the Anderson-Darling test statistic

Figure 4 P-values and values of the Anderson-Darling test statistic performed for years 1993–2016

One can observe that the p-values of neither of the tests performed to test the hypothesis (1) attain values greater than the level of significance, hence in all considered years we reject the hypothesis (1). On the other hand all of the performed tests suggest, that the best fit for each year of the studied period provides the lognormal distribution.

We may notice in Figure 4a that the p-values of the Anderson-Darling test, when testing the hypothesis (3) for the years from 1993 to 2005, are close to the level of significance $\alpha = 0.05$. Furthermore, it can be observed in Figure 4b that some of the values of the Anderson-Darling test statistic in this time period lie above the values of the critical function in the given year. A closer look at the p-values of the Anderson-Darling test of the hypothesis (3) and the values of the respective test statistic and the critical values up to year 2000 are collected in Table 1. The year 2000 is the last one when the hypothesis (3) was rejected. It is clear from Table 1 that the hypothesis (3)

year	1993	1994	1995	1996	1997	1998	1999	2000
p-value	0.0227	0.0408	0.0586	0.0420	0.0568	0.0538	0.0342	0.0490
test static	3.1631	2.6627	2.3637	2.6384	2.3894	2.4335	2.8130	2.5113
critical value	2.4941	2.4941	2.4941	2.4940	2.4940	2.4941	2.4941	2.4940

Table 1 Anderson-Darling test results for years from 1993 to 2000, years when the hypothesis (3) is rejected are in red

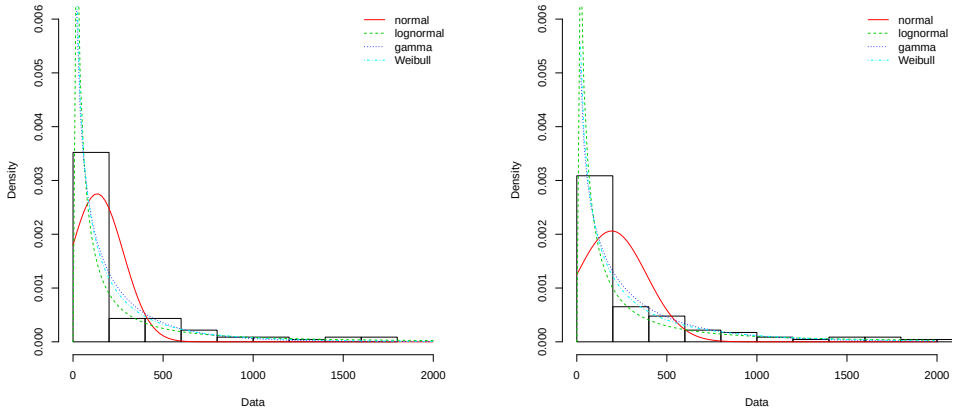
tested by the Anderson-Darling test is rejected for years 1993, 1994, 1996, 1999 and 2000 (denoted by red).

It should be noted that in general the p-value of the Kolmogorov-Smirnov test, when testing a given hypothesis for a given year, is greater than the p-value of the Cramer-von Mises test for the corresponding hypothesis and year, which is in turn greater than the p-value of the Anderson-Darling test (Figures 2a, 3a and 4a). We thus conclude that the Anderson-Darling test appears to be the most strict one for this dataset.

In all Figures 2, 3 and 4 one can notice that the fit of the gamma and Weibull distributions is initially poor for years from 1993 to 2005, but then significantly improves in years 2006–2008 to be almost on par with the fit of the lognormal distribution. This suggests that some change in the distribution of military expenditure has happened between years 2005 and 2008. When comparing the empirical densities in the respective years (see Figure 5) it

probabilities	0.01	0.05	0.10	0.90	0.95	0.99
quantiles – normal	−222.7324	−103.7562	−40.3305	407.1395	470.5653	589.5415
quantiles – lognormal	2.1000	6.7226	12.5004	994.1412	1848.5634	5917.7717
quantiles – gamma	0.1698	2.3652	7.4221	604.7310	834.7128	1390.1139
quantiles – Weibull	0.2975	2.9827	8.2556	647.5655	939.5561	1725.9847
empirical – quantiles	2.5909	5.5824	9.5225	736.3058	1287.5864	1974.7292

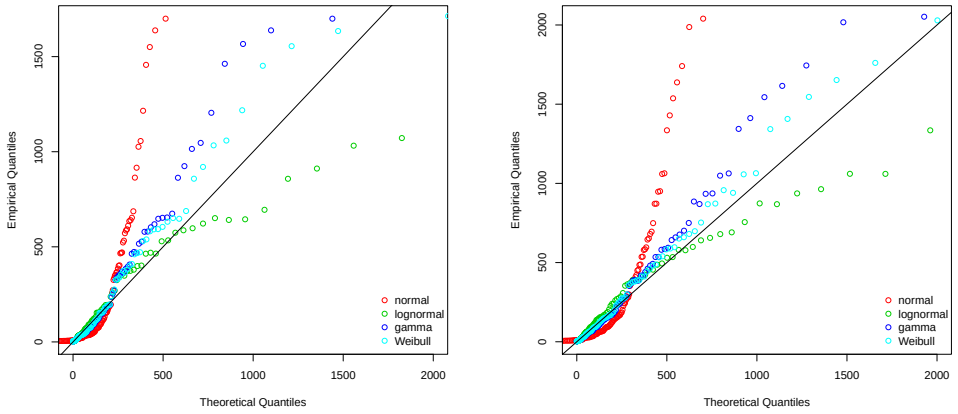
Table 2 Selected quantiles of the fitted distributions and the empirical quantiles in 2016



(a) Comparison of theoretical and empirical densities of distributions for year 2005

(b) Comparison of theoretical and empirical densities of distributions for year 2008

Figure 5 Comparison of theoretical and empirical densities of normal, lognormal, gamma and Weibull distributions



(a) Q-Q plot for year 2005

(b) Q-Q plot for year 2008

Figure 6 Q-Q plots of theoretical quantiles of normal, lognormal, gamma and Weibull distribution and empirical quantiles of the sample

may be noticed that military expenditure of countries experienced a slight increase in years between 2005 and 2008, most noticeably among countries, whose spendings were rather small prior to the year 2005. This is further supported by the Q-Q plots depicted in Figure 6.

Table 2 contains quantiles of estimated distributions and empirical quantiles of military expenditure in 2016. It was shown in the previous part of the contribution that the normal distribution is not an acceptable model for analyzed data. We come to the same conclusion when comparing the quantiles of the normal distribution with empirical quantiles or quantiles of other studied distributions. Hence, the quantiles of the normal distribution are also included in the table 2. One can observe from this comparison the mismatch between the empirical and the normal distribution at its tails. This further support the point that normal distribution is not suitable for modeling the data. If one analyzes quantiles of the gamma and Weibull distribution, they appear to be more or less comparable. Quantiles of the lognormal distribution are substantially different. They are close to the empirical quantiles for small probabilities, but their values for probabilities 0.95 and 0.99 are considerably greater. Estimates of quantiles for high probabilities can be taken as a marker of potential security risk. If military expenditure exceeds given value (say 95% quantile), one can deduce that military spendings in such a country are large (extreme) compared to the rest of the world. The choice of proper probability distribution strongly influences these estimates.

The ability to accurately describe the probability distribution of military expenditure, especially its high values, is useful from a security point of view. The empirical distribution suggests that military spending is expected to exceed \$1287.6 in 5% of countries. The corresponding quantiles of estimated distributions differ significantly

from this value. For example, 95% quantile of the log-normal distribution has a value of $1845.6, 95% quantile of Weibull distribution is $939.6. The description of the probability distribution of military expenditure using selected (parametric) models is not accurate enough in the area of higher values. A possible solution to this problem can be the use of nonparametric estimation methods, a topic we would like to focus on in the future.

5 Conclusion

The topic of this contribution was to find a suitable probability distribution of military expenditure per capita from 1993 to 2016. The statistical analysis involves data of 115 countries. The authors focus on a normal, lognormal, gamma and Weibull distribution. The quality of a distribution fit is based on comparison of an empirical and a theoretical cumulative distribution function and was tested by Anderson-Darling, Cramer-von Mises and Kolmogorov-Smirnov test. The distribution of the analyzed variable is not symmetric, so the normal distribution was rejected by all tests. According to the test results, the best fit is obtained by using the lognormal model. A more detailed view of the quantiles showed that there are differences between the probability distributions. The models based on the gamma and Weibull distribution were similar in terms of estimated quantiles. Quantiles of the lognormal distribution model are substantially different. They are close to the empirical quantiles for small probabilities (1%, 5% quantiles), but their values for 0.95 and 0.99 are considerably greater. The lognormal distribution seems to be a reasonable model for military expenditure per head. However, it should be noted that the ability of this probability distribution model to describe extreme values (large values of military spending indicating a potential security risk) is not sufficient.

Acknowledgements

Supported by the grant No. SV17-FVL-K102-AMB of the Ministry of Education, Youth and Sports, Czech Republic.

References

[1] Anděl, J. (2011). *Základy matematické statistiky.* Ed. 3. Praha: Matfyzpress.

[2] Bartošová, J. (2006). Rozdělení příjmů a jeho modely. In *8. národní konference Statistické dny v Brně* (pp. 1–10). Brno: Česká statistická společnost.

[3] Busschaert, P., Geeraerd, A., Uyttendaele, M. & VanImpe, J. (2010). Estimating distributions out of qualitative and (semi)quantitative microbiological contamination data for use in risk assessment. *International Journal of Food Microbiology*, 138, 260–269.

[4] D'Agostino, R. B. & Stephens, M. A., (1986). *Goodness-of-Fit Techniques*. New York: Marcel Dekker.

[5] Delignette-Muller, M. L. & Christophe Dutang (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64(4), 1–34. Available from: http://www.jstatsoft.org/v64/i04/

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

[7] Faraway, J., Marsaglia, G, Marsaglia J. & Adrian Baddeley (2017). goftest: Classical Goodness-of-Fit Tests for Univariate Distributions. R package version 1.1-1. Available from: https://CRAN.R-project.org/package=goftest

[8] Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis* [online]. 51(2), (pp. 904–917) [cit. 2018-04-19]. ISSN 0167-9473. Available at: https://doi.org/10.1016/j.csda.2005.09.011

[9] Goodness of fit tests in R. (2015) *R-bloggers: R news and tutorials contributed by R bloggers.* [online]. [cit. 2017-03-25] Available from: https://www.r-bloggers.com/goodness-of-fit-test-in-r/

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

[11] Střelec, L. (2015) Notes on power of normality tests of error terms in regression models. In SIMOS, T E. (Eds.), *Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2014 (ICNAAM 2014).* (pp. 7200131–7200134). Melville, New York: American Institute of Physics.

[12] Ul Hassan, M. & Pr Stockhammar, P. (2016). Fitting probability distributions to economic growth: a maximum likelihood approach, *Journal of Applied Statistics*, 43(9), 1583–1603.

Comparison of several modern numerical methods for option pricing

Tomáš Tichý[1], Jiří Hozman[2], Michal Holčapek[3], Dana Černá[4], Aleš Kresta[5]

Abstract. Option pricing is a popular problem of financial mathematics and optimization due to the non-linearity in the option pay-off function and enormous sensitivity to the selection of underlying processes and input parameters. This aspect differentiates options from other derivatives. Since pricing and hedging of plain vanilla options under the conditions of Gaussian distribution (or a so called Black-Scholes model) is already well documented, it commonly serves as a benchmark for developing of new approaches and methods, which, in fact, aims on options with more complex payoffs (exotic options) and/or probability distributions that fit empirical observations about the market prices better, but for which no analytical formula is available. Obviously, being able to compare the results of the novel model with theoretically correct one is a crucial step of model testing. In this contribution we focus on numerical pricing of options. We first review well known approaches and subsequently we analyze three novel approaches, discontinuous Galerkin approach, wavelet approach and fuzzy transform technique. Extensive comparative study for various input data and pay-off functions is provided.

Keywords: option pricing, Black-Scholes model, discontinuous Galerkin method, wavelet method, fuzzy transform

JEL classification: C44
AMS classification: 90C15

1 Introduction

Option pricing is a complex problem of financial mathematics. Due to the non-linearity in the option pay-off function and enormous sensitivity to the selection of underlying processes and input parameters, static hedging portfolio cannot be in general constructed (this aspect differentiates options from other derivatives). Instead, it is inevitable to create a dynamic portfolio strategy utilizing no-arbitrage arguments (for a review, see, e.g., [3]).

Since pricing and hedging of plain vanilla options under the conditions of Gaussian distribution (or a so called Black-Scholes (BS) model, [1] and [6]) is already well documented, it commonly serves as a benchmark for developing of new approaches and methods, which, in fact, aims on options with more complex payoffs (exotic options) and/or probability distributions that fit empirical observations of market prices better, but for which no analytical formula is available.

Obviously, being able to compare the results of the novel model with theoretically correct one is a crucial step of model testing. In this paper we extend our previous research on the topic of numerical pricing of options ([5, 8]), and specifically we study the impact of the moneyness (i.e., ITM, OTM and ATM options are considered) on the pricing error within the BS setting. The findings might show us the directions of further development of all analyzed methods.

We proceed as follows. After introductory section we briefly summarize the foundations of option pricing and selected novel numerical approaches. The core results of the paper are provided in Section 4.

2 Problem of option pricing

Since (European) option gives its holder a right to make a trade with the underlying asset x at maturity time T, i.e., in contrary to forwards or swaps there is no obligation for the holder, the option valuation requires more or less advance approaches. The fair price of an option, in dependency on the underlying (including its return volatility σ) and maturity, can be specified, utilizing the no-arbitrage arguments, by a so called Black-Scholes-Merton partial

[1]Department of Finance, VSB-TU Ostrava, Sokolska 33, 701 21, Ostrava, Czech Republic, tomas.tichy@vsb.cz
[2]Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic, jiri.hozman@tul.cz
[3]Institute for Research and Applications of Fuzzy Modeling, NSC IT4Innovations, University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic, michal.holcapek@osu.cz
[4]Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic, dana.cerna@tul.cz
[5]VSB-TU Ostrava, Sokolska 33, 701 21, Ostrava, Czech Republic, ales.kresta@vsb.cz

differential equation with riskless rate r as follows:

$$\frac{\partial u}{\partial t} + \mathcal{L}_{BSM}(u) = 0 \text{ in } Q_T = \Omega \times (0,T),\ \Omega \subset \mathbb{R}^+ \tag{1}$$

with

$$\mathcal{L}_{BSM}(u) = -\frac{1}{2}\sigma^2 x^2 \frac{\partial^2 u}{\partial x^2} - rx\frac{\partial u}{\partial x} + ru. \tag{2}$$

This formulation is known according to the authors who assumed options with simple payoff functions (plain vanilla options) underlying asset of which obeys geometric Brownian motion.

By solving the equation above accompanied by proper terminal (or initial) condition given by the payoff function and suitable boundary conditions related to limiting values of the option, we can get no-arbitrage solution, i.e. fair market price of any option. However, the system leads to analytical solution only under specif conditions. Thus, since the terminal condition can be complex and/or the underlying process not easy dealt with, it might be necessary to apply some numerical approximation.

3 Novel numerical methods

In this section, three novel numerical approaches, namely Discontinuous Galerkin, Wavelet, Fuzzy Transform, are briefly summarized.

Discontinuous Galerkin Method

Discontinuous Galerkin (DG) method, in some way, combines the methods of finite volumes and finite elements. For the option price $u(x,t)$ it provides discrete solution U_l^{DG} on $t_l = l\tau$ with constant time step $\tau = T/M$ on complete computation domain Ω, i.e.

$$U_l^{DG} \in S_h \equiv \{v \in L^2(\Omega);\ v|_{I_k} \in P_p(I_k)\ \forall I_k \in \mathcal{T}_h\} \tag{3}$$

where $\mathcal{T}_h$ split Ω and $P_p(I_k)$ denotes the space of all polynomials of order less or equal to p defined on subintervals of I_k with length h, see [4].

Wavelet Methods

Wavelet methods (WM) are complex adaptive methods that allow efficient solution of differential equations of various types. The method consists in a representation of the solution U_l^W on the l-th time level in a wavelet basis Ψ, i.e.

$$U_l^{WM} = \sum_{\psi_\lambda \in \Psi} u_\lambda \psi_\lambda. \tag{4}$$

Due to the properties of a wavelet basis Ψ, many coefficients u_λ are small and they can be thresholded and the solution is represented by the small number of parameters. In numerical experiments we used a wavelet basis from [2].

Fuzzy Transform Method

The fuzzy transform (FT) method has been proposed in [7] as a novel approximation technique based on tools of fuzzy modeling. The core of fuzzy transform consists in a fuzzy partition of the subspace Ω by means of a system of fuzzy sets $\Phi = \{\phi_k(x); k \in \mathbb{K}, x \in \Omega\}$ with $\mathbb{K} \subseteq \mathbb{N}$ such that

$$\sum_{k \in \mathbb{K}} \phi_k(x) = 1, \quad x \in \Omega. \tag{5}$$

The fuzzy sets $\phi_k(x)$ of a fuzzy partition of Ω are called the basic functions. Using the basic functions $\phi_k(x)$ the functions of $L^2(\Omega)$ space are transformed onto finite vectors of real numbers that provide compressed information about original functions. The components of real vectors are called the components of fuzzy transform. An approximation of original functions of $L^2(\Omega)$ space is then obtained by the linear combination of components of a fuzzy transform and basic functions of a fuzzy partition of Ω. In order to solve partial differential equations, the components of fuzzy transform technique are used instead of function values in more standard techniques.

4 Numerical comparison

The numerical analysis presented below is based on data originally used by the authors ([4]) when pricing vanilla put options at German option market on September 15, 2011 using DG approach. Here, only short maturity options

(74 calendar days) on DAX (German stock market index) are considered – given that the index value was 4,700, the strike prices of 3,900, 4,700, and 5,500 indicate three different kinds of moneyness; in particular, OTM, ATM, and ITM option, respectively.

The volatility used for pricing of options is implied by the market prices of options, which means that the price obtained using analytical BS formula exactly matches the market prices. Moreover, since the market provide both values (BS price, implied volatility), we can estimate relevant risk free rate as 3.9% per annum easily.

Numerical approximation is crucially related to the discretization of the domain; here, we set the maximal underlying asset price as 8 times the strike price (it should help us to offset the impact of the boundary condition) and the time step as $1/3600$. Each of the methods is considered in the form of linear as well as non-linear (quadratic, cosine) approximation.

In the case of vanilla options, the quality of the approximation can be easily obtained by comparing the results with the theoretical price according to the BS model. We can consider relative error in the L^2-norm on the whole computational domain as follows: $e_{L^2} = \frac{||u_M^* - u(T)||}{||u(T)||}$, where u_M^* denotes the approximate solution obtained by one of three presented numerical approaches.

The results of corresponding errors are apparent from Figure 1 for OTM, ATM and ITM option (from top to bottom). In all cases we can observe rather stable decrease of the error with increasing number of basis functions. It is quite apparent that quadratic approximation provides much better results, except for FT approach for which the differences are apparent only for lower number of basis functions.

When comparing particular methods, the lower error is obtained when Wavelet approach is considered, while the worst results are provided by FT approach. However, it can be given by the novelty of the approach and relatively low knowledge how to optimize its setting. If we focus on differences between OTM, ATM, and ITM options, we can hardly see any – DG approach shows some differences for ATM options, when quadratic approximation provides slightly lower error. These findings can be confirmed if additional levels of moneyness are considered and generally also for options with varying maturity.

5 Conclusion

This paper presents selected convergence results of three (relatively) novel approaches to numerical pricing of options, namely Discontinuous Galerkin method, Wavelet methods, and Fuzzy Transform approach. For the comparison analysis, short term OTM, ATM, and ITM options (i.e., options with different level of moneyness) on German stock market index have been selected. The results do not show any apparent differences when moneyness is varying – in all cases Wavelet methods provide the best results (in terms of specified error function), while Fuzzy Transform approach the worst. The future research should therefore be focused on possible parameter optimization of the latter approach.

Acknowledgements

Supported by the project No. 16-09541S of the Czech Science Foundation (GACR) and SGS project of VSB-TU Ostrava SP2018/34.

References

[1] Black, F., Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *J. Political Econ.*, 81, 637–659.

[2] Černá, D., Finěk, V. (2011). Construction of Optimally Conditioned Cubic Spline Wavelets on the Interval. *Adv. Comput. Math.*, 34, 219–252.

[3] Delbaen, F., Schachermayer, W. (2006). *The Mathematics of Arbitrage*. Berlin: Springer.

[4] Hozman, J, Tichý, T. (2014). Black-Scholes option pricing model: Comparison of h-convergence of DG method with respect to boundary condition treatment. *ECON – Journal of Economics, Management and Business*, 24(4), 141–152.

[5] Hozman, J., Černá, D., Holčapek, M., Tichý, T., Valášek, R. (2018). Moderní numerické metody při oceňování opcí: Porovnání na případu jednoduchých vanilla opcí. *Politická ekonomie* (submitted).

[6] Merton, R.C. (1973). Theory of rational option pricing. *Bell J. Econ.*, 4, 141–18.

[7] Perfilieva, I. (2003). Fuzzy transform: Application to the reef growth problem. In *Fuzzy Logic in Geology*, R. Demicco and G. Klir, Eds. Amsterdam: Academic Press, chap. 9, 275–300.

[8] Tichý, T., Hozman, J., Holčapek, M. (2017). A note on several alternatives to numerical pricing of options. In *Liberec Economic Forum* (pp. 381–189), Liberec: Technical University of Liberec.

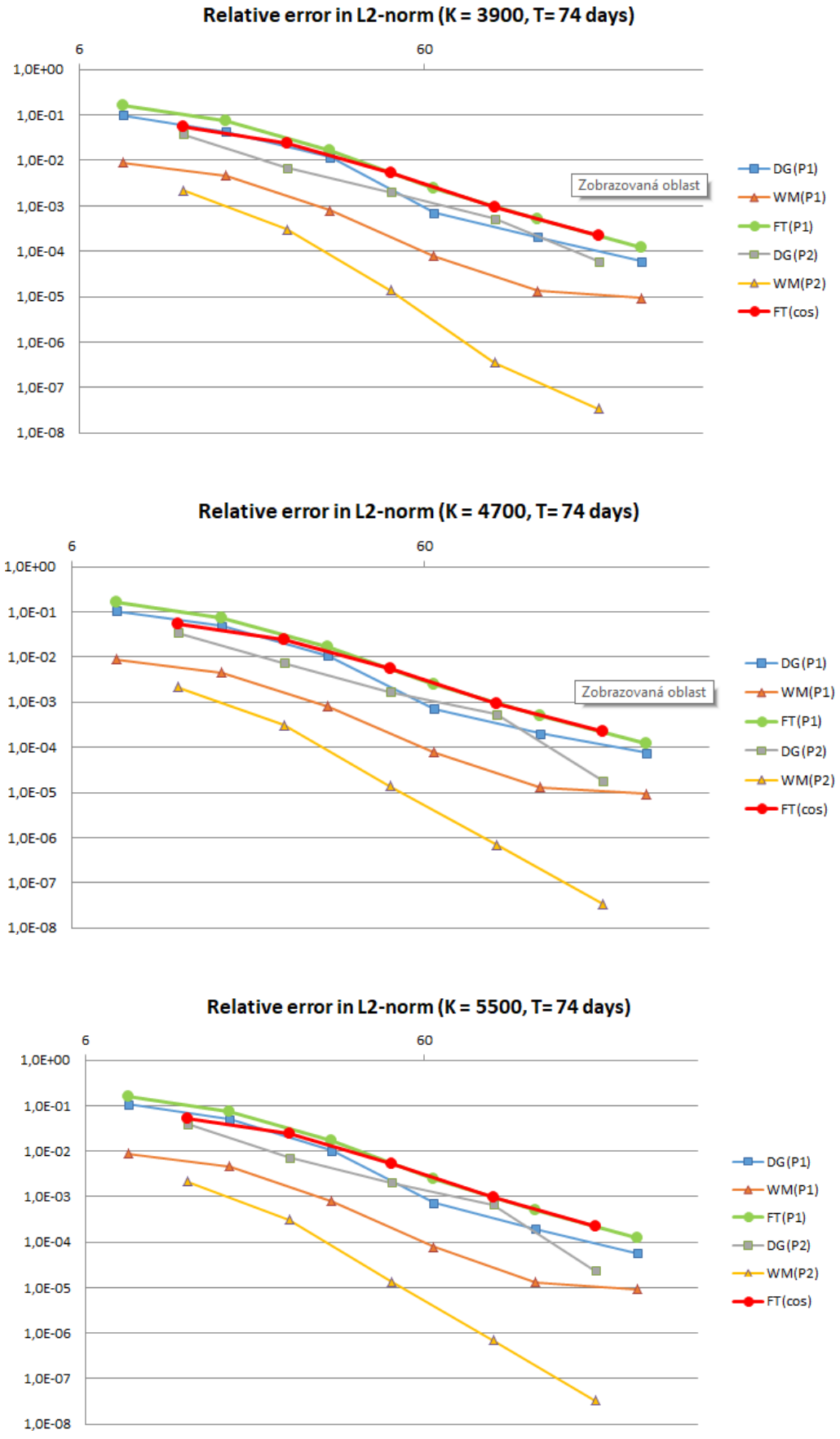


Figure 1 Relative error of option price for OTM (top), ATM (middle), and ITM (bottom) options

Comparison of decision variants with reference set (BIPOLAR method) in selection process of suppliers

Agnieszka Tłuczak[1]

Abstract. Bipolar is one of the Multiple Criteria Decision Analysis (MCDA) methods, proposed by Ewa Konarzewska-Gubała [6, 7, 8]. The essence of the analysis in the BIPOLAR method consists in a fact that alternatives are not compared directly to each other, but they are confronted to the two sets of reference objects: good and bad. Some alternatives can be evaluated as over good, i.e. better than at least one of desirable reference object or underbid, i.e. worse than at least of one nonacceptable object.

Proposed method takes into account all stages of the decision-making process, starting with the selection of criteria and determining their weights, by modelling the preferences of decision-makers, on obtaining recommendations in the form of a ranking of decision-making variants. In this paper case study of choosing suppliers is presented. The BIPOLAR method has been used to select suppliers of food products for retail stores.

Keywords: BIPOLAR, multi-criteria decision making, reference system.

JEL Classification: C44

1 Introduction

In practice, most decision problems can be described using three variables: the objective, decision variants and usability of the solution. In uncomplicated decision problems, the objective function depends on the usefulness, in the sense of the optimal result, estimated on the basis of the adopted assessment criteria. In the case of a large number of evaluation criteria, in principle, it is not possible to choose an optimal, but only suboptimal decision. None of the decisions will meet all the assessment criteria better than the others. In addition, the decision-maker's preferences, sometimes unequivocally indicating the accepted solution, play an important role in complex decision-making problems. In logistics, tasks of multi-criteria decision making are very common. The problem of choosing (eg. a supplier) or an investment decision (eg. in the area of linear / point expansion of the logistics infrastructure) has been and remains the most common problem in the functioning of individual enterprises and entire supply chains. Each decision has certain consequences (sometimes far-reaching). This fact makes logistic decisions a very complex problem. In order to avoid mistakes and randomness, it is necessary to refer to scientific methods that have been tested in practice. Such methods include multi-criteria decision making. They are opposed to a single-criterion analysis because they try to express a cohesive family of criteria as an instrument of understandable, acceptable and exhaustive communication. This should enable creation, justification and transformation of preferences in the decision-making process [5].

The multi criteria decision making/analysis is a field of study originating from operations research, which aims the development of mathematical procedures and advanced computer-based-methods that support the decision maker in solving multiple criteria decision problems. Today, these methods are gaining in popularity and are increasingly used in decision-making process. This article will present BIPOLAR method, which allows choosing the best logistics provider that deals with deliveries. In the Bipolar method elements of ELECTRE methodology [12] and ideas of Merighi (1980) [11] algorithms of confrontation can be found. The method has been used in applications (Jakubowicz [3], Jakubowicz and Konarzewska-Gubała [4], Dominiak [1,2], Konarzewska-Gubała [9, 11]). Moreover, the method has also been applied to model multi-stage multi-criteria decision processes [13, 14, 15].

2 BIPOLAR method

The bipolar method is a multicriteria decision support method; it allows sorting and organizing decision variants. Decision variants are compared with a bipolar reference set that contains "good" and "bad" objects. We assume that a finite set of decision variants $A=\{a^1, a^2, \ldots, a^m\}$ and a set of criteria $F=\{f1, f2, \ldots, f_n\}$, $f_k:A\rightarrow K_k$, for $k=1, 2, \ldots, n$ where K_k are given by a numerical, ordinal or binary scale. the criteria are defined in such a way that higher ratings of pre-test criteria are in relation to lower ratings [7, 14]. The decision-maker assigns weight w_k to each criterion, where $\sum_{k=1}^{n} w_k = 1$ and $w_k \geq 0$ for $k=1, 2, \ldots, n$, equivalence threshold q_k and veto threshold v_k. The decision maker also gives the minimum level of compliance of the criterion assessments s as the concordance level. We assume that the condition $min\{w_j\} \leq s \leq 1$ is fulfilled [9, 10, 15].

The decision maker sets a bipolar reference system $R=D \cup Z$ which is composed of good $D=\{d^1,d^2,...,d^d\}$ and bad reference objects $Z=\{z^1, z^2,..., z^z\}$. We assume that $D \cap Z = \emptyset$. The number of elements of the set R is equal to $d+z$. Elements of set R are denote as r^j, $j=1, 2,...,d+z$ and $\forall d \in D \forall z \in Z \forall k = 1, ..., n\ f_k(d) \geq f_k(z)$.
The Bipolar method consists of three phases. The first phase is a comparison of the decision alternative with the reference objects. This allows for the determination of prevalence indices and the structure of preferences in the reference system. In phase two, the position of the decision variant relative to the bipolar reference system should be determined. Third phase involves reasoning about the relations in the set of decision variants.
Ph.1 For the pair (a^i, r^j), where $a^i \in A$ and $r^j \in R$ according to the formula [7, 8, 9, 14, 15]:

$$c^+(a^i, r^j) = \sum_{k=1}^{n} w_k \varphi_k^+(a^i, r^j), \quad (1)$$

$$c^-(a^i, r^j) = \sum_{k=1}^{n} w_k \varphi_k^-(a^i, r^j), \quad (2)$$

$$c^=(a^i, r^j) = \sum_{k=1}^{n} w_k \varphi_k^=(a^i, r^j), \quad (3)$$

all values are calculated, where

$$\varphi_k^+(a^i, r^j) = \begin{cases} 1, & if\ f_k(a^i) - f_k(r^j) > q_k \\ 0, & otherwise \end{cases}, \quad (4)$$

$$\varphi_k^-(a^i, r^j) = \begin{cases} 1, & if\ f_k(r^j) - f_k(a^i) > q_k \\ 0, & otherwise \end{cases}, \quad (5)$$

$$\varphi_k^=(a^i, r^j) = \begin{cases} 1, & if\ \left|f_k(r^j) - f_k(a^i)\right| > q_k \\ 0, & otherwise \end{cases}, \quad (6)$$

Set of indices I^+ and I^- are defined as follow [14]:

$$I^+(a^i, r^j) = \{k: \varphi_k^+(a^i, r^j) = 1\} \text{ and } I^-(a^i, r^j) = \{k: \varphi_k^-(a^i, r^j) = 1\}. \quad (7)$$

The outranking indicators $d^+(a^i, r^j)$and $d^-(a^i, r^j)$ are given by formulas [8]:
a) if $c^+(a^i, r^j) > c^-(a^i, r^j)$ and the non-discordance test: $\forall k \in I^- f_k(a^i) > v_k$ is positively verified, then

$$d^+(a^i, r^j) = c^+(a^i, r^j) + c^=(a^i, r^j) \wedge d^-(a^i, r^j) = 0 \quad (8)$$

b) if the non-discordance test isn't positively verified then $d^+(a^i, r^j) = 0$ and $d^-(a^i, r^j) = 0$
a) if $c^+(a^i, r^j) < c^-(a^i, r^j)$ and the non-discordance test:

$$\forall k \in I^+ f_k(a^i) > v_k \quad (9)$$

is positively verified, then

$$d^+(a^i, r^j) = 0 \text{ and } d^-(a^i, r^j) = c^-(a^i, r^j) + c^=(a^i, r^j) \quad (10)$$

b) if the non-discordance test isn't positively verified then $d^+(a^i, r^j) = 0$ and $d^-(a^i, r^j) = 0$
a) if $c^+(a^i, r^j) = c^-(a^i, r^j)$ and the non-discordance test:

$$\forall k \in I^+ f_k(a^i) > v_k \quad (11)$$

is positively verified, then

$$d^+(a^i, r^j) = c^+(a^i, r^j) + c^-(a^i, r^j) \text{ and } d^-(a^i, r^j) = c^-(a^i, r^j) + c^+(a^i, r^j) \quad (12)$$

b) if the non-discordance test isn't positively verified then $d^+(a^i, r^j) = 0$ and $d^-(a^i, r^j) = 0$

This outranking indicators allow to determine the relations: large preference L, indifference I and incomparability R, which are defined as follows [14]:

$$a^i L r^j, if\ d^+(a^i, r^j) > s \wedge d^-(a^i, r^j) = 0 \quad (13)$$

$$r^j L a^i, if\ d^+(a^i, r^j) = 0 \wedge d^-(a^i, r^j) > s \quad (14)$$

$$a^i L r^j, if\ d^+(a^i, r^j) > s \wedge d^-(a^i, r^j) > s \quad (15)$$

$$a^i R r^j, otherwise \quad (16)$$

Ph.2 Each variant is compared with the set of good and bad objects, for each variant a^i the success achievement degree d_{iS} is determined [15]:

$$d_{iS} = \begin{cases} d_{iD}^+ = \max\{d_{ih}^+: a^i L d^h \vee a^i I d^h\}, if\ \{h: a^i L d^h \vee a^i I d^h\} \neq \emptyset \\ d_{iD}^- = \min d_{ih}^-, \text{if } \{h: a^i L d^h \vee a^i I d^h\} \neq \emptyset \wedge \{h: d^h L a^i\} \neq \emptyset \\ d_{iD}^+ = d_{iD}^- = 0\ otrherwise \end{cases} \quad (17)$$

The failure avoidance degree d_{iN} is given by formula:

$$d_{iN} = \begin{cases} d_{iZ}^- = \max\{d_{ik}^-: z^k L a^i \vee z^k I a^i\}, if\ \{k: z^k L a^i \vee z^k I: a^i\} \neq \emptyset \\ d_{iZ}^+ = \max, if\ \{k: z^k L a^i \vee z^k I a^i\} \neq \emptyset \wedge \{k: a^i L z^k\} \neq \emptyset \\ d_{iD}^+ = d_{iD}^- = 0\ otrherwise \end{cases} \quad (18)$$

The values d_{iS} determine the position of the variant a^i relative to the D set, and d_{iN} determine the position of the variant a^i relative to the Z set.
Ph.3 Each variant is describe by vector $[d_{iS}, d_{iN}]$, these values are needed to sort and order variants.
According to success achievement degree d_{iS} values allow sort variants on three categories [8, 9]:

S_1 categorie: a^i , $d_{iS}=d_{iD}^+>0$
S_2 catgorie: a^i , $d_{iS}=d_{iD}^-<0$
S_3 categorie: a^i , $d_{iS}=0$
Ignoring from the analysis variants in the third category and assuming that each category variant of S_1 is submitted over any variant of category S_2, we order linear variants in the following way :
a^1 is better than a^2:
$(d_{1S}=d_{1D}^+ \wedge d_{2s}=d_{2D}^+ \wedge d_{1D}^+ > d_{2D}^+) \vee (d_{1S}=d_{1D}^+ \wedge d_{2s}=d_{2D}^-)$
$\vee (d_{1S}=d_{1D}^- \wedge d_{2s}=d_{2D}^- \wedge d_{1D}^- < d_{2D}^-)$ (19)
a^1 is equivalent to a^2:
$(d_{1S}=d_{1D}^+ \wedge d_{2s}=d_{2D}^+ \wedge d_{1D}^+ = d_{2D}^+) \vee (d_{1S}=d_{1D}^- \wedge d_{2s}=d_{2D}^- \wedge d_{1D}^- = d_{2D}^-)$ (20)
Due to failure avoidance degree, the d_{iN} values allow you to sort decision variants into three categories:
N_1 categorie: a^i: $d_{iN}=d_{iZ}^+ > 0$
N_2 categorie: a^i: $d_{iN}=d_{iZ}^- > 0$
N_3 categorie: a^i: $d_{iN}=0$
Ignoring the analysis of objects of category N_3 and assuming that each variant of category N_1 is preferred over any variant of category N_2, the linear order of objects is made in the following way [9, 10]:
a^1 is preferred to a^2:
$(d_{1N}=d_{1Z}^+ \wedge d_{2N}=d_{2N}^+ \wedge d_{1Z}^+ > d_{2Z}^+) \vee (d_{1N}=d_{1Z}^+ \wedge d_{2N}=d_{2Z}^-) \vee (d_{1N}=d_{1Z}^- \wedge d_{2N}=d_{2Z}^- \wedge d_{1Z}^- < d_{2Z}^-)$ (21)
a^1 is equivalent to a^2: $(d_{1N}=d_{1Z}^+ \wedge d_{2N}=d_{2Z}^+ \wedge d_{1Z}^+ = d_{2Z}^+) \vee (d_{1N}=d_{1Z}^- \wedge d_{2N}=d_{2Z}^- \wedge d_{1Z}^- = d_{2Z}^-)$ (22)
Considering the problem of the joint assessment of success achievement degree and the assessment of the degree of failure avoidance degree, we define three categories of variants:
B_1 categorie: a^i: $d_{iD}^+>0 \wedge d_{iZ}^+ > 0$
B_2 categorie: a^i: $d_{iD}^->0 \wedge d_{iZ}^+ > 0$
B_3 categorie: a^i: $d_{iD}^->0 \wedge d_{iZ}^- > 0$
assuming that each variant of category B_1 is preferred over any variant of category B_2, and any variant of category B_2 is preferred over each variant of category B_3 we sort linearly the variants in each class as follows:
B_1:
a^1 is preferred to a^2: $(d_{1S}+d_{1N})>(d_{2S}+d_{2N})$
a^1 is equivalent to a^2: $(d_{1S}+d_{1N})=(d_{2S}+d_{2N})$
B_2:
a^1 is preferred to a^2: $(1-d_{1S}+d_{1N})>(1-d_{2S}+d_{2N})$
a^1 is indifferent to a^2: $(1-d_{1S}+d_{1N})=(1-d_{2S}+d_{2N})$
B_3:
a^1 is preferred to a^2: $(d_{1S}+d_{1N})<(1-d_{2S}+d_{2N})$
a^1 is equivalent to a^2: $(d_{1S}+d_{1N})=(d_{2S}+d_{2N})$

3 Case study: application of BIPOLAR method in supplier selection

We are considering a discrete problem of multi-criteria decision making, in which three suppliers are evaluated on the basis of two real-valued criteria: f_1- price, f_2 - reliability of supply. Suppliers are rated on a scale from 0 to 10 for both criteria, where 0 - the worst rating, 10 - the best rating. The bipolar reference system consists of two good objects - $D = \{d_1, d_2\}$ and one bad objects - $Z = \{z_1\}$. The evaluation of decision variants and reference objects in relation to the criteria is presented in Table 1 and Figure 1.

	f_1	f_2
a_1	8	6
a_2	5	9
a_3	6	8
d_1	8	8
d_2	9	9
z_1	3	3

Table 1. Values of criteria for decision variants and reference objects.

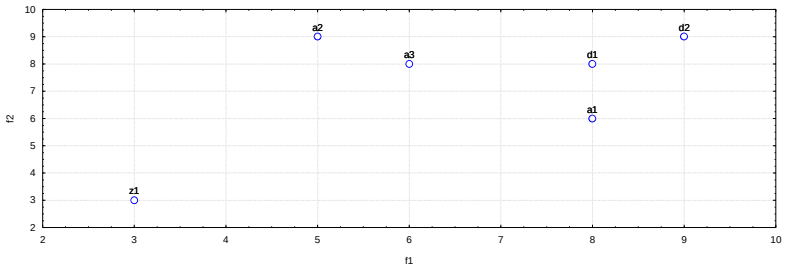


Figure 1. Graphical illustration of the criteria values for decision variants and reference objects.

As a result of the analysis of the decision-maker's preferences, it was assumed, that veto threshold is*:* v_1=0, v_2=0, additionally weights are: w_1=0,45; w_2=0,55, equivalence threshold are q_1=0,30, q_2=0,30 and concordance level s=0,6.

According to formula [4]-[6] for $k=1, 2$ and all pairs of decision variants and reference objects values of $\varphi_1^+(a^i,r^j), \varphi_1^-(a^i,r^j), \varphi_1^=(a^i,r^j)$ were calculated. All values have been used to find coefficient C^+, C^-, $C^=$ (tab. 2) which are given by formula [1]-[3].

C^+	d_1	d_2	z_1	C^-	d_1	d_2	z_1	$C^=$	d_1	d_2	z_1
a_1	0	0	1	a_1	0,55	1	0	a_1	0,45	0	0
a_2	0,55	0	1	a_2	0,45	0,45	0	a_2	0	0,55	0
a_3	0	0	1	a_3	0,45	1	0	a_3	0,55	0	0

Table 2. Matrix of coefficient C^+, C^-, $C^=$

Next step it was to determine the matrix of outranking indicators D^+ and D^- (tab. 3):

D^+	d_1	d_2	z_1	D^-	d_1	d_2	z_1
a_1	0	0	1	a_1	1	1	0
a_2	0,55	0	1	a_2	0	1	0
a_3	0	0	1	a_3	1	1	0

Table 3. Matrix of outranking indicators D^+ and D^-.

This outranking indicators allow to determine the relations: L, I and R, which are defined as follows (tab.4) :

	d_1	d_2	z_1
a_1	a_1Rd_1	a_2Rz_1	a_1Lz_1
a_2	a_2Ld_1	a_1Rd_2	a_3Rz_1
a_3	a_2Rd_1	a_3Rd_2	a_3Lz_1

Table 4. Relations defining the structure of preferences.

Next the position of the decision variant against the bipolar reference set was determined according to the formula [17-18] to determine the success achievement degree and to failure avoidance degree (tab. 5).

	S	N
a_1	0	1
a_2	0,55	0
a_3	0	1

Table 5. The success achievement degree and to failure avoidance degree

Due to the success achievement degree and avoiding failure decision's variants have been sorted into appropriate categories (tab. 6).

mono ranking due to success achievement degree	**a_2** **a_1, a_3; a_1 is equivalent to a_3**
mono ranking due to failure avoidance degree	a_2 a_1, a_3; a_1 is equivalent to a_3

Table 6. Mono-ranking

Considering jointly evaluation of success achievement degree and failure avoidance degree, the following categories of variants are given:

$a_2 \in B_1$

$a_1, a_3 \notin B_2 \wedge a_1, a_3 \notin B_3$.
Bipolar ranking looks like: 1. a_2, and variants $a_1$and a_3 are equivalent decision variants.

4 Summary

Among many methods that can help in choosing the supplier, the BIPOLAR method should be indicated. Based on the research presented in this article, it can be concluded that BIPOLAR method is proper to choose. The analyzes carried out were aimed at selecting one among three suppliers. The proposed procedure allows comparing the considered variants with "good" objects and "bad: objects. As a consequence, the decision variants were ranked considering the jointly evaluation of success achievement and failure avoidance.

References

[1] Dominiak C. (1997). Portfolio Selection Using the Idea of Reference Solution. In: G. Fandel, T. Gal (eds.) *Multiple Criteria Decision Making*. Springer-Verlag.
[2] Dominiak C. (2006). Application of modified Bipolar method. In: T. Trzaskalik (ed.) *Multicriteria Methods on Polish Financial Market*, PWE (in Polish).
[3] Jakubowicz S. (1987). *Work Characteristics of a „Good" Physics Teacher on the Basis of His Lessons*. RPBP.III.30.VI.4.6. The University of Wrocław (in Polish).
[4] Jakubowicz, S., Konarzewska-Gubała, E. (1989). *Work Characteristics of a Physics Teacher*. University of Wrocław (copied manuscript, in Polish).
[5] Kauf S., Tłuczak A. (2016). *Optymalizacja decyzji logistycznych*. Warszawa: Difin.
[6] Konarzewska-Gubała E. (2002). Multiple Criteria Company Benchmarking Using the BIPOLAR Method. In T.Trzaskalik, J.Michnik (eds.) *Multiple Objective and Goal Programming. Recent Developments.* Physica-Verlag. Springer-Verlag.
[7] Konarzewska-Gubała E. (1989). *BIPOLAR: Multiple Criteria Decision Aid Using Bipolar Reference System*, LAMSADE, Cahier et Documents no 56, Paris.
[8] Konarzewska-Gubała E. (1987). *Multicriteria Decision Analysis with Bipolar Reference System: Theoretical Model and Computer Implementation*. Archiwum Automatyk i Telemechaniki, vol. 32, no 4.
[9] Konarzewska-Gubała, E. (1991). *Multiple Criteria Decision Aid: System Bipolar*. Scientific Works of the University of Economics in Wrocław, no 551 (in Polish).
Konarzewska-Gubała, E., 1996. *Supporting an effective performance appraisal system*. Argumenta Oeconomica vol.1.
[10] Merighi D.(1980). *Un modello di valutazione rispetto insiemi di riferimento assegnati*. Ricerca Operativa no 13.
[11] Roy B.(1985). *Methodologie Multicritere d'Aide a la Decision*. Economica, Paris.
Trzaskalik T.(1987). Model of multistage multicriteria decision processes applying reference sets. In: *Decision Models with Incomplete Information,* Scientific Works of the University of Economics in Wrocław, no 413, (in Polish).
[12] Trzaskalik T., Sitarz S. (2012). How to Deal with Overgood and Underbad Alternatives in Bipolar Method. In: Watada J., Watanabe T., Phillips-Wren G., Howlett R., Jain L. (eds) I*ntelligent Decision Technologies. Smart Innovation, Systems and Technologies*, vol 16. Springer, Berlin, Heidelberg
[13] Trzaskalik T., Sitarz S. (2007). Underbad and Overgood Alternatives in Bipolar Method, In: Zadnik Stirn L., Drobne S. (Eds), *Proceedings of the 9th International symposium on Operational research in Slovenia 2007*.
[14] Trzaskalik T., Sitarz S., Dominiak C. (2013). Unified procedure for bipolar method, in: Zadnik Stiretod, In: Zadnik Stirn L., Zerovnik J., Povh J., Drobne S., Lisec A. (Eds), *Proceedings of the 12th International Symposium on Operational Research in Slovenia 2013*.

A Semiparametric Approach to Modeling Time-Varying Quantiles

Petra Tomanová[1]

Abstract. A conditional quantile of a time series can be simply estimated by inverting the (possibly time-varying) conditional distribution. However, it turns out that a parametric assumption about the underlying distribution is felt to be too restrictive. We focus on distribution-free filters for time-varying quantiles. Our approach is motivated by the fact that the first derivatives of the quantile criterion function are, in a sense, similar to conditional score. We propose a Beta-t-IG model for time-varying quantile estimation based of Generalized Autoregressive Score (GAS) framework, also known as Dynamic Conditional Score (DCS) models, allowing parameters to vary over time and capturing the dynamics of time-varying parameters by the autoregressive term and the scaled score of the conditional observation density. We compare our model to the originally proposed specifications of CAViaR models.

Keywords: Generalized autoregressive score model, quantile regression, time-varying quantile.

JEL classification: C14, C22
AMS classification: 91G70

1 Observation-driven models for time-varying quantiles

In this paper, we focus on the observation-driven models, where the time variation of quantiles is introduced by letting quantiles be functions of lagged dependent variables as well as lagged exogenous variables. Thus the quantiles are perfectly predictable given the past information. Our approach to time-varying quantile modeling is inspired by the generalized autoregressive score (GAS) models of [1], also known as Dynamic Conditional Score (DCS) models of [4], which are described in Section 1.1. Then we focus on the semiparametric models and propose a new model for value-at-risk (VaR) estimation in the Section 1.2. Next Section 2 shows the model performance using the real data and the Section 3 concludes.

1.1 Generalized autoregressive score models

In a general setting, let $\boldsymbol{y}$ be an observed time series of interest, $y_1, \ldots, y_n$, and $\tilde{\boldsymbol{f}}$ be an unobserved path of equal length $\tilde{f}_1, \ldots, \tilde{f}_n$. Assume that the data generating process is $\boldsymbol{y} \sim p(\boldsymbol{y}; \tilde{\boldsymbol{f}})$ where $\tilde{\boldsymbol{f}}$ represents a time-varying feature of the 'true' model density.

In an observation-driven model, the time-varying effect is extracted as a direct function of past data: $f_t = f_t(y_1, \ldots, y_{t-1}; \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a parameter vector. Then we have predictive model of the form

$$p(y_t | f_t, y_1, \ldots, y_{t-1}; \boldsymbol{\beta}). \tag{1}$$

Under GAS frameworks, for f_t we adopt the updating scheme

$$f_{t+1} = \beta_0 + \sum_{i=1}^{p} \beta_i f_{t-i+1} + \sum_{j=1}^{q} \beta_{p+j} s_{t-j+1}, \tag{2}$$

where s_t is an appropriate function of past data, $s_t = s_t(y_1, \ldots, y_t; f_1, \ldots, f_t; \boldsymbol{\beta})$. Let $s_t = S_t \nabla_t$, where

$$\nabla_t = \frac{\partial \ln p(y_t | f_t, y_1, \ldots, y_{t-1}; \boldsymbol{\beta})}{\partial f_t},$$

and S_t can be defined as $\mathcal{I}_{t|t-1}^{-1}$, $\mathcal{I}_{t|t-1} = E_{t-1}\left[\nabla_t \nabla_t'\right]$.

Once the time-varying parameters of the conditional distribution are estimated, we can simply obtain a conditional quantile of a time series by inverting the conditional distribution at each time t. However, it turns out that a parametric assumption about the underlying distribution is felt to be too restrictive. Thus, we focus on distribution-free filters for time-varying quantiles.

[1]University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czech Republic, petra.tomanova@vse.cz

1.2 Semiparametric models

In the general dynamic quantile regression problem, the (1) is specified as

$$y_t = f_t(\boldsymbol{x_{t-1}}; \boldsymbol{\beta}) + u_t, \tag{3}$$

where u_t is an unknown error term with a generally unspecified distribution, and $\boldsymbol{x_{t-1}}$ might contain a set of lagged values of the response y_{t-k}, $k > 0$, and explanatory variables. Let $f_t(\boldsymbol{x_{t-1}}; \boldsymbol{\beta}_\theta)$ be the conditional quantile at probability level θ, where $\boldsymbol{\beta}_\theta$ is the solution to

$$\min_{\boldsymbol{\beta}} \sum_t \rho_\theta(y_t - f_t(\boldsymbol{x_{t-1}}; \boldsymbol{\beta}_\theta)). \tag{4}$$

Function $\rho(\cdot)$ is a loss function, usually specified as $\rho(u) = u(\theta - I(u < 0))$. Function $f_t(\cdot)$ defines the dynamic link between the response y_t and the explanatory information $\boldsymbol{x_{t-1}}$.

If the observations are assumed to come from an asymmetric double exponential distribution, maximising the log-likelihood function is equivalent to minimising the criterion function in (4) [2]. The Skewed-Laplace location-scale family $(SL(\mu, \tau, \theta))$ of distributions has density function

$$p_\theta(u) = \frac{\theta(1-\theta)}{\tau} \exp\left\{-\rho_\theta\left(\frac{u-\mu}{\tau}\right)\right\}, \tag{5}$$

where μ is the mode and $\tau > 0$ is a scale parameter. Let u_t be an i.i.d. sequence in the model (3), $u_t \sim SL(0, \tau, \theta)$, then the likelihood function becomes

$$L_\theta(\boldsymbol{\beta}, \tau; \boldsymbol{y}, \boldsymbol{x}_{t-1}) \propto \tau^{-n} \exp\left\{-\tau^{-1} \sum_{t=1}^{n} (y_t - f_t(\boldsymbol{x_{t-1}}; \boldsymbol{\beta}))[\theta - I(y_t - f_t(\boldsymbol{x_{t-1}}; \boldsymbol{\beta}) < 0)]\right\}, \tag{6}$$

where $I(\cdot)$ is an indicator function.

The maximum likelihood estimate for $\boldsymbol{\beta}$ in (6) is equivalent to the quantile estimator in (4) since (4) is contained in the exponent of the likelihood. De Rossi and Harvey [2] pointed out that quantiles are fitted separately, hence there is no notion of an overall model for the whole distribution and assuming the distribution (5) for one quantile is not compatible with assuming it for another. Setting up this particular parametric model is simply a convenient device that leads to an appropriate criterion function for what is essentially a nonparametric estimator. Now we can derive a basic score function for updating equation (2).

Differentiating (minus) (4) at all points where this is possible gives

$$\sum_t IQ(y_t - f_t(\boldsymbol{x_{t-1}}; \boldsymbol{\beta}_\theta)),$$

where

$$IQ(y_t - f_t(\boldsymbol{x_{t-1}}; \boldsymbol{\beta}_\theta)) = \begin{cases} \theta - 1, & \text{if } y_t < f_t(\boldsymbol{\beta}_\theta, \boldsymbol{x_{t-1}}) \\ \theta, & \text{if } y_t > f_t(\boldsymbol{\beta}_\theta, \boldsymbol{x_{t-1}}) \end{cases}$$

defines the *quantile indicator*. Since $\rho(\cdot)$ is not differentiable at zero, $IQ(0)$ is not determined [2]. Then the basic model for time-varying quantiles becomes

$$\begin{aligned} y_t &= f_t + u_t \\ f_{t+1} &= \beta_0 + \beta_1 f_t + \beta_2 s_t \\ s_t &= IQ(y_t - f_t) \end{aligned}$$

where we write $f_t = f_t(\boldsymbol{x_{t-1}}; \boldsymbol{\beta})$ for simplicity. By maximizing likelihood function (6) we obtain estimate $\boldsymbol{\beta}_\theta$ and then conditional quantiles $f_t, t = 2, \ldots, n$, can be computed. It is important to emphasize that, though we treat (6) exactly as a likelihood function, methods here do not actually assume that the observations $\boldsymbol{y}$ follow a Skewed-Laplace distribution. The likelihood form here is only employed because it leads to a mathematically equivalent estimator to (4).

We call this model as a *Quantic* model (based on [2], where authors used it for quantile indicator IQ). De Rossi and Harvey [2] argue that only indicator variables should be used in the context of nonlinear dynamic models for conditional quantiles since specifications, which are based on actual values, rather than indicators, may suffer from a lack of robustness to additive outliers. Thus, we will examine this model performance in the Section 2.

CAViaR models

Engle and Manganelli [3] were among the first to use quantile regression and defined a general class of nonlinear dynamic models called the Conditional Autoregressive Value-at-Risk (CAViaR). They proposed four specifications of quantile f_t, see [3], which we compare to our specifications in the Section 2.

Mitrodima and Oberoi [9] pointed out, that the potential of CAViaR models in their various specifications has not been fully explored and thus, they propose two sets of CAViaR specifications in their paper. The first is a component structure to generate a long-range dependence and the second is to incorporate lower frequency returns as quantile predictors. They also presented a modified estimation algorithm that makes the model parameters estimation easier. Huang et al. [5] extend the CAViaR model by allowing the parameters to be a function of past returns. Schaumburg [11] adopts the Indirect GARCH CAViaR model to incorporate the previous day's return as a regressor in the quantile equation and proposes the Indirect Autoregressive Threshold GARCH model. Kuester et al. [7] incorporate autocorrelation in the return process. Jeon and Taylor [6] use information in the implied volatility of options to enhance CAViaR forecasts, both directly using a forecast combination strategy and by using the implied volatility as a regressor in the quantile specification. Rubia and Sanchis-Marco [10] consider a first-order autoregressive covariate-extended CAViaR model. De Rossi and Harvey [2] combine CAViaR with signal extraction. We show how it is possible to use the GAS models to derive models for time-varying quantile estimation.

Beta-t-IG model

The Beta-t-GARCH(1,1) model was introduced by [1] and [4] and belongs to the GAS class of models. For a sequence of financial returns $\{y_t\}_{t\in N}$ with time-varying conditional volatility and leverage effects the Beta-t-GARCH(1,1) model is defined

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, \\ \sigma_{t+1}^2 &= \omega + \beta\sigma_t^2 + (\alpha + \gamma d_t)\frac{(v+1)y_t^2}{(v-2) + y_t^2/\sigma_t^2}, \end{aligned}$$

where $\{\varepsilon_t\}_{t\in Z}$ is an i.i.d. sequence of standard Student's t random variables with $v > 2$ degrees of freedom and d_t is a dummy variable that takes value $d_t = 1$ for $y_t \leq 0$ and $d_t = 0$ otherwise.

If z_θ is the θ-quantile of the standard Student's t distribution, then $f_t = \sigma_t z_\theta$, where f_t is the θ-quantile of y_t. This implies

$$f_{t+1}^2 = \omega z_\theta^2 + \beta f_t^2 + (\alpha + \gamma d_t)\frac{(v+1)y_t^2 z_\theta^2}{(v-2) + y_t^2 z_\theta^2/f_t^2},$$

which is the Beta-t-IG version of the CAViaR model. By substitution $\beta_0 = \omega z_\theta^2$, $\beta_1 = \beta$, $\beta_2 = \alpha$ and $\beta_3 = \gamma$, we obtain the following updating equation for the quantile f_t

$$f_{t+1} = \left[\beta_0 + \beta_1 f_t^2 + (\beta_2 + \beta_3 d_t)\frac{(v+1)y_t^2 z_\theta^2}{(v-2) + y_t^2 z_\theta^2/f_t^2}\right]^{1/2}. \tag{7}$$

As [3] point out, though the original CAViaR models are exactly specified counterparts of corresponding GARCH models, CAViaR models are in fact more general than the GARCH models.

2 Empirical results

To compare our Beta-t-IG model for quantiles to the originally proposed specifications, we replicate results of [3]. Engle and Manganelli [3] used historical series of portfolio returns of General Motors (GM), IBM and S&P 500. We took the same sample of 3,392 daily prices from Datastream and computed the daily returns as 100 times the difference of the log of the prices. The samples ranges from April 7, 1986, to April 7, 1999. The first 2,892 observations are used to estimate the model and the last 500 are left for out-of-sample testing. We also repeat the analyses using new data sample ranging from November 18, 2010, to April 25, 2018, resulting in 1,639 in-sample observations and 300 out-of-sample observations.

We estimated 1% and 5% 1-day VaRs, using the four CAViaR specifications (Adaptive, IG, Symmetric absolute value (SAV) and Asymmetric slope (AS)) described in [3], as well as our Beta-t-IG and Quantic specification. For the adaptive model we set $G = 10$, where G entered the definition of the adaptive model, see [3], and for the Beta-t-IG model (7) we set $v = 10$. As Engle and Manganelli [3] noted, parameter G itself could be estimated (analogously v as well), however, this would go against the spirit of this model, which is simplicity. We optimized the models using the similar procedure as in [3]. The 5% VaR estimates for GM (original data sample) are plotted in Figure 1, and results the for Beta-t-IG and Quantic specifications are reported in Table 1.

	April 7, 1986 – April 7, 1999						November 18, 2010 – April 25, 2018					
	Quantic			Beta-t-IG			Quantic			Beta-t-IG		
	GM	IBM	S&P 500	GM	IBM	S&P 500	GM	IBM	S&P 500	GM	IBM	S&P 500
1% VaR												
β_0	**3.1880**	0.5327	**2.4661**	**2.4186**	0.1391	**0.3433**	**0.9596**	**8.4375**	**0.3029**	6.7110	1.6756	**0.2401**
standard errors	0.3036	0.4364	0.1783	1.4443	0.1351	0.0916	0.5695	0.7074	0.0385	4.6165	1.0979	0.1045
p values	0.0000	0.1111	0.0000	0.0470	0.1515	0.0001	0.0460	0.0000	0.0000	0.0730	0.0635	0.0108
β_1	**0.3282**	**0.8736**	**0.1348**	**0.6778**	**0.9467**	**0.8231**	**0.7907**	**-1.0045**	**0.8715**	0.3981	**0.6865**	**0.8422**
standard errors	0.0506	0.0989	0.0319	0.1468	0.0302	0.0380	0.1215	0.0024	0.0139	0.3393	0.1402	0.0897
p values	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1203	0.0000	0.0000
β_2	**-19.1226**	-6.2775	**-19.8512**	0.1140	-0.0038	-0.0079	-3.0385	0.0045	**-3.1150**	0.4500	0.1351	-0.0579
standard errors	0.8453	4.9461	0.6168	0.0815	0.0133	0.0239	2.6041	0.5799	0.2709	0.5250	0.2038	0.0618
p values	0.0000	0.1022	0.0000	0.0809	0.3877	0.3708	0.1216	0.4969	0.0000	0.1957	0.2537	0.1745
β_3				**0.1310**	**0.0991**	**0.2543**				-0.2165	0.0741	**0.3925**
standard errors				0.0764	0.0355	0.0698				0.4832	0.2622	0.0497
p values				0.0433	0.0026	0.0001				0.3270	0.3888	0.0000
RQ	167.09	175.43	105.08	168.65	177.93	104.99	99.04	85.82	45.50	93.94	83.63	46.42
Hits in-sample (%)	0.9336	1.0373	0.7607	1.0028	0.9682	0.9682	1.3423	0.7932	1.0982	1.0372	1.0982	0.9152
Hits out-of-sample (%)	1.2000	1.6000	1.6000	1.0000	1.8000	1.8000	0.0000	39.3333	0.6667	0.0000	2.0000	1.3333
DQ in-sample (p values)	0.6418	0.7637	0.0000	0.6532	0.7913	0.9566	0.5477	0.9943	0.0000	0.9629	0.9360	0.2632
DQ out-of-sample (p values)	0.9948	0.8310	0.0464	0.9994	0.0343	0.0518	0.0000	0.0000	0.9984	1.0000	0.5460	0.0002
5 % VaR												
β_0	**0.0319**	**0.2403**	**0.0087**	**0.1349**	**0.1076**	**0.0332**	**1.1608**	**0.1816**	**0.0251**	**1.3756**	0.1507	**0.1179**
standard errors	0.0076	0.0652	0.0022	0.0612	0.0588	0.0130	0.5372	0.0845	0.0079	0.4762	0.1169	0.0294
p values	0.0000	0.0001	0.0000	0.0137	0.0338	0.0054	0.0154	0.0157	0.0007	0.0019	0.0987	0.0000
β_1	**0.9872**	**0.9006**	**0.9916**	**0.9375**	**0.9029**	**0.8859**	**0.5964**	**0.9052**	**0.9782**	**0.6218**	**0.8078**	**0.8095**
standard errors	0.0030	0.0244	0.0018	0.0242	0.0218	0.0235	0.1847	0.0433	0.0055	0.0929	0.0744	0.0436
p values	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000
β_2	**-0.3282**	**-1.4844**	**-0.2304**	0.0102	**0.0248**	0.0190	**-2.2713**	**-0.9429**	**-0.5433**	0.1025	0.0676	-0.0213
standard errors	0.0496	0.3389	0.0456	0.0126	0.0145	0.0155	0.5435	0.3739	0.1129	0.0885	0.0524	0.0341
p values	0.0000	0.0000	0.0000	0.2104	0.0432	0.1105	0.0000	0.0058	0.0000	0.1233	0.0986	0.2662
β_3				**0.0512**	**0.0835**	**0.1531**				0.1529	**0.1417**	**0.3338**
standard errors				0.0182	0.0271	0.0316				0.0951	0.0729	0.1022
p values				0.0025	0.0010	0.0000				0.0540	0.0260	0.0005
RQ	550.08	526.88	308.97	548.38	516.48	300.78	331.13	232.38	171.00	329.47	230.82	164.16
Hits in-sample (%)	5.3942	5.0138	5.7746	4.9447	4.9447	4.9793	4.8810	4.5149	5.8572	5.0031	5.0641	4.9420
Hits out-of-sample (%)	6.0000	7.4000	7.0000	5.8000	6.6000	6.6000	4.0000	4.3333	5.3333	5.6667	5.3333	3.3333
DQ in-sample (p values)	0.7342	0.6060	0.0391	0.6827	0.7404	0.7065	0.8243	0.7983	0.3089	0.8677	0.6600	0.2335
DQ out-of-sample (p values)	0.7995	0.0083	0.0553	0.7754	0.0992	0.0004	0.7899	0.5486	0.4743	0.6508	0.0606	0.0006

Table 1 Estimates and statistics for Quantic and Beta-t-IG specification. Significant coefficients at 5% confidence level are formatted in bold.

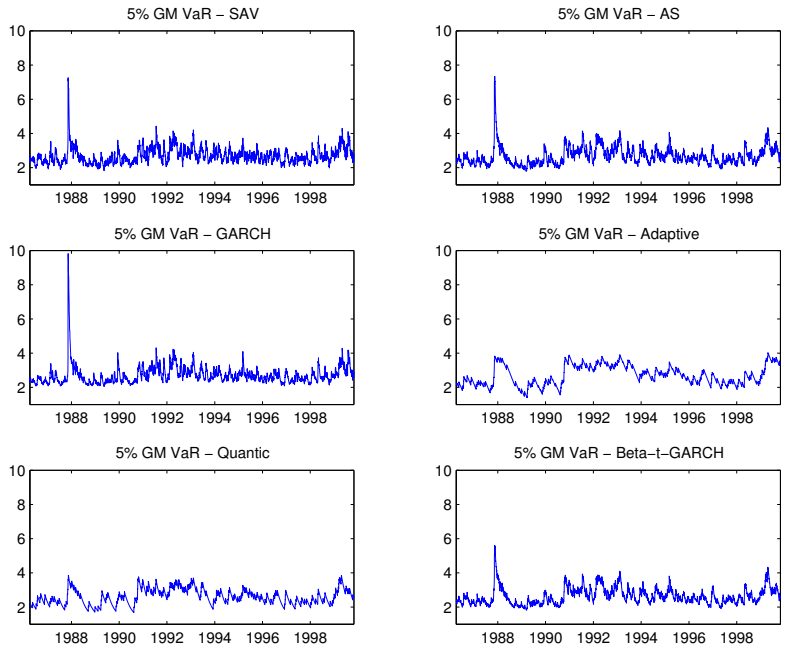


Figure 1 Estimated 5% VaR_t, where $VaR_t = -f_t(\boldsymbol{x_{t-1}}; \boldsymbol{\beta}_\theta)$.

Harvey [4] noted, that the three specifications (IG, SAV, AS) may suffer from a lack of robustness to additive outliers, which is clear from the Figure 1 (the spike at the beginning of the sample which is caused by the 1987 crash). The issue might be solved by (i) using only indicator variables as in the Adaptive or the Quantic specification, or by (ii) using a weighting scheme to put smaller weights to the extreme observations. In the Beta-t-IG specification, the updates are less responsive to extreme realized returns than in the standard IG model based on the Gaussian weighting scheme. This makes the computed quantiles less responsive to abrupt volatility changes. In contrast, if there are incidental tail observations, the Beta-t-IG model provides a much better and more robust estimate of the quantile at time t [8]. Figure 2 plots the impact curve for the 5% VaR estimates of S&P 500. The sharp difference between the impact of positive and negative returns in the Beta-t-IG and the AS model suggests that there might be relevant asymmetries in the behavior of the 5% quantile of this portfolio.

Table 1 presents the values of estimated parameters, corresponding standard errors and (one-sided) p values, value of the regression quantile objective function (4), the percentage of times the VaR is exceeded, and the p value of the Dynamic Quantile (DQ) test, both in-sample and out-of-sample in the same manner as in [3].[2] The Beta-t-IG and Quantic models are comparably precise as the originally proposed specification, as measured by the percentage of in-sample hits, however, the Quantic model has a slightly inferior performance for 5% VaR. On average, the Beta-t-IG model provides the best fit in terms of the RQ value among all 6 models and very good performance in terms of both in-sample and out-of-sample DQ test, especially for the GM. The Beta-t-IG model performs poorly for S&P 500 out-of-sample when estimating 5% VaR. However, Engle and Manganelli [3] documented that all their models fail to predict both 1% and 5% VaR for S&P 500 in terms of out-of-sample DQ test at the 5% confidence level. They explained such results by the fact that the last part of the sample of the S&P 500 is characterized by a sudden spur of volatility and roughly coincides with our out-of-sample period. Finally, the coefficients β_3 of Beta-t-IG model are always significant at the 5% confidence level for the original data sample, which indicates the presence of strong asymmetric impacts on VaR of lagged returns. However, when analyzing the new data sample, results about asymmetric impact are rather mixed.

3 Conclusion

In this paper, we showed how it is possible to use the GAS models to derive models for time-varying quantiles. We proposed a new Beta-t-IG specification for VaR estimation and compared it to the originally proposed CAViaR specification of [3] and to specification based on indicator variables. Although Engle and Manganelli [3] documented that GARCH might provide an unsatisfactory approximation when applied to tail estimation since GARCH implicitly assumes that the tails follow the same process as the rest of the returns, we show that our Beta-t-IG solves

[2] See [3] for more information about computation and testing procedure.

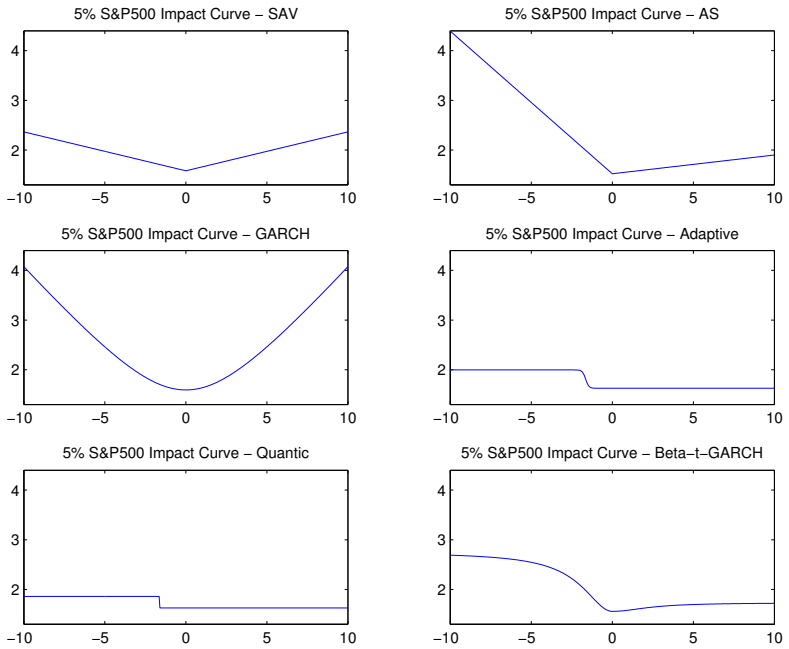


Figure 2 Impact Curve

this issue and performs well in terms of both in-sample and out-of-sample DQ test. In the future work, we will focus on generalization of this finding and propose a GAS-based framework to time-varying quantile modeling.

Acknowledgements

Supported by the grant No. F4/93/2017 of the Internal Grant Agency VŠE.

References

[1] Creal, D., Koopman, S. J. & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5), 777–795.

[2] De Rossi, G & Harvey, A. C. (2006). *Time-varying quantiles*. University of Cambridge, Faculty of Economics.

[3] Engle, R. F. & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4), 367–381.

[4] Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*. Cambridge University Press.

[5] Huang, D., Yu, B., Lu, Z., Fabozzi, F., J., Focardi, S. & Fukushima, M. (2010). Index-Exciting CAViaR: a new empirical time-varying risk model. *Studies in Nonlinear Dynamics & Econometrics*, 14(2).

[6] Jeon, J. & Taylor, W. (2013). Using CAViaR Models with Implied Volatility for Value-at-Risk Estimation. *Journal of Forecasting*, 32(1), 62–74.

[7] Kuester, K., Mittnik, S. & Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1), 53–89.

[8] Lucas, A. & Zhang, X. (2016). Score-driven exponentially weighted moving averages and Value-at-Risk forecasting. *International Journal of Forecasting*, 32(2), 293–302.

[9] Mitrodima, G. & Oberoi, J., S. (2017). Value at Risk models of autoregressive quantiles for improved performance on financial criteria.

[10] Rubia, A. & Sanchis-Marco, L. (2013). On downside risk predictability through liquidity and trading activity: A dynamic quantile approach. *International Journal of Forecasting*, 29(1), 202–219.

[11] Schaumburg, J. (2012). Predicting extreme value at risk: Nonparametric quantile regression with refinements from extreme value theory. *Computational Statistics & Data Analysis*, 56(12), 4081–4096.

Modeling of ESG factors influence on both long term risk management and return on investment

Oldřich Trenz [1], Edward Kassem [2], Sylvie Formánková[3], Jan Kolomazník[4], Oldřich Faldík[5]

Abstract. The paper deals with the modeling of ESG factors impacts on the long-term risks management in conjunction with the expected return on investment. ESG factors (environment, social responsibility and corporate governance), the current state of ESG investing, and the evaluation of sustainable investment are described. As an input data for individual assessments, both ESG investment funds and funds not containing these descriptive instruments are used. The selective portfolio theory (Selective Markowitz's Model) and multi-factor or multi-dimensional statistical methods are employed to evaluate the return on investment risk. The emphasis is placed on the social responsibility of the investment in financial funds. The achieved results could be used as recommendations of the investment decision making process as well as a basis of investment decision making software.

Keywords: Sustainable investing, SRI, ESG, Sustainability, Corporate social responsibility, Decision making

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

In the last few years, the investors are becoming more interested in investments in unit trusts which, in their investments, apply the rules of Social Responsible Investment (SRI), which is a part of Corporate Social Responsibility (CSR), or alternatively, ESG criteria are being applied – environmental, social and administrative responsibilities [17], [19]. The funds that apply the SRI investment form avoid investments into areas of tobacco and alcohol, weapon manufacturing, the running of gambling and other areas which are antagonistic to social responsibility. Contrary to this, the funds select investments that fulfill ESG criteria, while their portfolio contains areas which meet high standards in the field of environmental protection, ecology, employee handling ethics (1st world countries) and which apply transparent company management [5].

SRI funds are not only funds focused on green technologies, but they also have important representation [21]. These funds also respect the carbon footprint (Carbon Intensity) and environmental impact (Climetrics). Energy conservation, renewable resources, waste and water management are very enticing at present and according to BNP Paribas [2], 15 % of investments effectuated in supportive stimulative packages go into this branch, while this trend can be recently seen in China, where these "stimulants" account for 9 % of the total amount of carried out investments. Today, BNP Paribas offers direct SRI investments support and administers, in these funds, investments totalling around 32 billion EURO [2]. In the USA, the present approach is called "The Green New Deal" with emphasis placed on the support of sustainability and the prevention of economic crisis, whence, as part of the activation package, 106 billion USD, from the total of 787 billion USD, is allocated specifically to this activity, thus supporting socially responsible investments [18].

Some studies contradict the benefits of the SRI investments, i.e., they point out that the performance of the funds which do not fulfill the sustainable investment form is greater than the performance of funds that comply with SRI. Some go as far as to recommend diverging from the SRI [3]. At present, however, the character of investment and investment markets is changing into a form in which it will be easier to acquire capital means

[1] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, oldrich.trenz@mendelu.cz.
[2] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, Edward.kassem@mendelu.cz.
[3] Mendel University in Brno/ Faculty of Business and Economics/Department of Management, Zemědělská 1, 613 00 Brno, Czech Republic, sylvie.formankova@mendelu.cz.
[4] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic,jan.kolomaznik@mendelu.cz.
[5] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic,oldrich.faldik@mendelu.cz.

specifically for investing into SRI funds. This way, then, it will be possible to gain a certain competition advantage [4]. Forbes along with and talks with experts show that we can expect a better return of fund investments and ESG strategies [8]. Sonia Kowal, the president of Zevin Asset Management, a company managing 0,6 billion USD and focusing on ESG investments, says that companies with the best employee relationships and the best environmental sustainability normally show a better financial performance in the long-term, as opposed to companies on the other side of the spectrum [10].

The present situation, in the scope of the SRI investments, in Czech Republic, is fairly tenable. From the 10 largest european SRI fund managers, 4 representatives offer investments into these funds. Namely: BNP Paribas, the Belgian KBC (ČSOB), the French Société Generale (KB) and the Austrian Erste Group (ČS). These administrators manage approximately 20 % of the european market of SRI funds for small investors. The image of the SRI is an option or renewing trust in the market after the crash of capital markets which occurred in the midst of the period of several past years. Positive expectations are newly being manifested with small and middle-sized investors, in connection with putting forth potential investments. Representatives of investment funds that administer SRI investments are aware of this movement and they see a great investment potential for the future here [1].

2 Material and Methods

The investor, when investing, needs to assess the potential risk connected with the investment values and the overall impact of this enterprise. Sustainable, responsible, and impact investing (SRI) spans a wide range of products and asset classes, embracing not only public equity investments (stocks), but also cash, fixed income and alternative investments, such as private equity, venture capital and real estate [20].

A study performed in 2015 by the Oxford University and the Arabesque Partners analyzed over 200 sources (academic studies, industry reports, newspaper articles, book chapters and books), the result being that 88 % of the reviewed sources attained the result of a positive SRI impact on the investments, in the form of their higher performance [4].

Although the SRI fund portfolio may be, compared to other funds, smaller, the available data may be assessed from various viewpoints, i.e., the character of the SRI itself, as well as of the other funds, and this with respect to their risk-level, as well as their performance-level. Fund productivity is assessed always in the given time-interval; in the case of this contribution, we take into account funds' performance in the last 5 years, the last 3 years, the last year and the last six months. Performance, in this description, means the % increase/decrease of the fund in the given period. When assessing the rating (risk-level) of the fund, the situation is more complex. There are several methodologies which assess the quality of the given fund, while sometimes it is difficult to perform the conversion between the different evaluation approaches. For the purposes of assessing the fund data, the SRRI (synthetic risk and productivity indicator) methodology will be used – one that is used by, for instance, the Erste Group. The value of this indicator can appear between 1–7, whilst it connects risk and volatility. The limit value 1 is equal to small risk, the value of 3 to a medium risk and 7 corresponds to high risk [7].

The aim of the article is to compare funds with the ESG enhancement with other funds, and this at the level of their rating and performance, with the help of the portfolio theory [6]. Emphasis will be placed on verifying the approach (portfolio theory) comparatively with the statistical approach. This method having been verified, the outputs will be used for a web-based project designated for the support of small investors who are considering investments into instruments within the area of sustainable funds. It should duly be added that the objective of this article is not to fully plant the portfolio theory. Instead, there is the primary testing of this approach, its analysis, and furthermore a follow-up of more research and the consequent application.

The aim of the article is a familiarization with the issue of unit trusts, with the evaluation in the form of the SRI, optionally the ESG rating; and, a proposal of a simplified model for the assessment of potential investments into these funds, with emphasis on the diversification of the investment risk. The outputs of this study will be used for creating a web-application enabling the support of decision-making of small investors in this specific sector, with the option of the levelling of investments into rated and unrated funds, including the option of keeping the investment profile of the registered user.

2.1 Portfolio Theory – the Selective Markowitz Model

The portfolio theory – the selective Markowitz model – joins the expected portfolio output, portfolio risk, the effective boundary of the investment into the evaluated funds, and it enables the creation of an optimal investment portfolio – i.e., to diversify the investment risk. The fundamental premise of the model is that the investors are sensible and all of them invest for an equally long time-period, while the risk is given by the standard deviation [11].

We are dealing with a portfolio made up of *n* investment instruments with every instrument *R* having a certain weight *w*, this weight being such that the total sum is equal to 1. The weights in this case may even be negative, which means the given instrument must be shortened. The expected yield is then given by the formula:

$$E(R_P) = \sum_{i=1}^{n} w_i . R_i \tag{1}$$

Thus, the portfolio yield is a weighted average of the yields of the individual assets. The expected risk of the individual assets can be measured using the standard deviation of the yields. The expected risk-level of the portfolio is not only a weighted average, as it had been in the case of the expected yield. The standard deviation of the whole portfolio depends mainly on the co-variance of the yields of the individual assets. The following relationship shows the standard deviation of the portfolio with *n* activities:

$$\sigma_p = \sqrt{\sum_{i=1}^{n} \sum_{i=1}^{n} w_{i.} w_j . \sigma_{ij}} \tag{2}$$

where σ_p pis the expected standard deviation of the portfolio yields, w_i w_j are the wieghts of the asset *i* and *j* and σ_{ij} is the co-variance between the yields *i* and *j*.

If the individual assets are not perfectly correlated (correlation near 1), it will always be possible to find such a portfolio whose expected risk will be lower than the expected risk of the safest asset. The diversification effect does not necessarily have to depend on the number of assets that are included in the portfolio. What is important is the mutual low correlation between the assets [12]. A high mutual correlation can be seen, for instance, with the stocks from the same sector. There is a certain sector dependence. The yields of such stocks will react in a very similar fashion to the world's events, to the established macro-economic conditions in the region – i.e., they will not be, in view of the executed investments, stable. Thus, a portfolio composed of tens of titles of the same market segment will not necessarily bear the same fruits like a portfolio composed of several stocks from different sectors.

Markowitz's optimal portfolio is such a portfolio $X_{opt,}$ for which there is an indifference curve u_k such, that:

$$\{(R_{opt}, \sigma_{opt})\} = u_k \cap EM, \tag{3}$$

Where the set of acceptable portfolios is assessed and where EM is a set of effective portfolios in Markowitz's sense; for more details, see, e.g., Edwin [6].

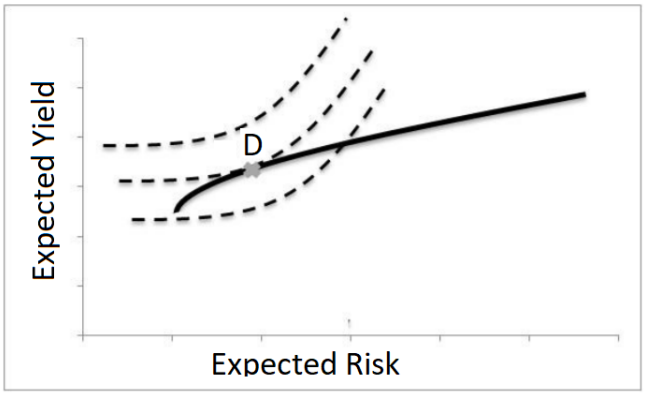


Figure 1 Optimal Portfolio of the Individual Investor

The optimal portfolio is the intersection between the effective boundary and the indifference curve of the investor, which is a tangent to this boundary (it is, therefore, located top-left). The optimum portfolio is demonstrated in Figure 1. Portfolio D is optimal for the investor, because it lies on the highest positioned indifference curve which still has a common point with the portfolio.

For the calculation itself, the environment of the Python programming language and the freely available source codes of the Markowitz method tailored for this programming language will be used [15].

2.2 Statistical Methods Used – Pearson's Chi-Squared Test

Pearson's chi-squared test is the basic and most widely used pair test of independence in a contingent chart. The zero hypothesis, in this approach, is the statement that stochastic magnitudes X_1 and X_2 are independent; the alternative hypothesis is then the mutual dependence of the stochastic magnitudes. Therefore, there is a higher probability that the stochastic magnitude X_1 will not affect the stochastic magnitude X_2, in the case of the zero-hypothesis [16],

$$\chi_P^2 = \sum_{i=1}^{R} \sum_{j=1}^{S} \frac{(n_{ij} - m_{ij})^2}{m_{ij}}, \tag{4}$$

where the statistical chi-squared reaches values from the interval $<0; n.(q-1)>$, where $q = min\{R, S\}$. The test is based on the comparison of the observed frequency (experiment) and the so-called expected frequencies (based on the validity of the H_0 hypothesis) of the distinct combinations of the stochastic magnitudes X_1 and X_2 [13].

The chi-squared independence test does not, however, say anything about the power of the relationship; it only does not dismiss, or dismisses, the zero hypothesis of the dependence of the stochastic magnitudes (signs) X_1 and X_2. For the identification of the power of the relationship, it is useful to employ an adequate statistical coefficient. The tested data will be assessed in the Statistica software.

2.3 Input Data

The input data designated for verifying the approach has been the data of capital funds with ESG-respecting factors; data available on the pages of the Socially Responsible Investment[6]. The advantage of this search engine is the option of filtering the funds based on the appurtenance to a given country where this fund is marketable. The data used has been filtered for the Czech Republic and a partial view of this data is available in Table 1.

Fund id	**Rating**	**Performance**			
		6m	1yr	3yr	5yr
LU0132414144	4	-6.71	-3.47	1.26	3.70
LU0963865323	4	-2.24	-6.35	-1.75	*2.10*
LU0616241476	3	-0.48	2.45	0.08	3.06
LU0119099819	3	-1.65	1.23	1.10	2.53
LU0518421895	3	-0.01	2,00	0.17	3.38
...	...	...	...	...	...

Table 1 ESG Funds Data

Data which does not take into account ESG factors in financial instruments has been acquired in a similar way, Table 2. From available databases of funds and their descriptive information, the Morning Star[7] database has been used. Here, detailed data of the performance of individual funds is available.

Fund id	**Rating**	**Performance**			
		6m	1yr	3yr	5yr
LU0329443377	4	4,18	4.67	0.69	13.95
LU0613078186	4	2.37	6.40	-3.50	-6.38
LU0557862082	4	3.07	8.01	2.73	3.72
LU0329445158	4	4.89	11.17	11.94	17.66
LU0551246555	2	-0.56	-0.28	1.42	2.87
...	...	...	...	...	...

Table 2 Other Funds Data (non-ESG)

[6] Aocially Responsible Investment – https://yoursri.com/.
[7] Morning Star – http://www.morningstar.co.uk/uk/.

For the collection of data from the given sources, we have used a mechanical processing, i.e., in the form of the designated script. A manual processing, vis á vis the number of source data needed and their interconnectivity, does not seem to be a good and a tenable solution in terms of time.

3 Results

Based on the application of the Markowitz model on select investment funds of the ESG and other funds, results presented in Figure 1 have been reached. It stands to reason that for achieving better evidence, it is necessary to input a larger data set of assessed data at the start. The data hereby used (from tens of funds) do not fully reflect the described reality in the given funds area, and it would be suitable to evaluate the funds across a whole area or by the investment focus. In the volatility values achieved with ESG funds there is an evident fluctuation of the fund value; this fact, then, points to a higher risk, i.e., to a loss of executed investments. This finding is not a happy one in relation to ESG funds, but some studies suggest that it is not out of the question, [4].

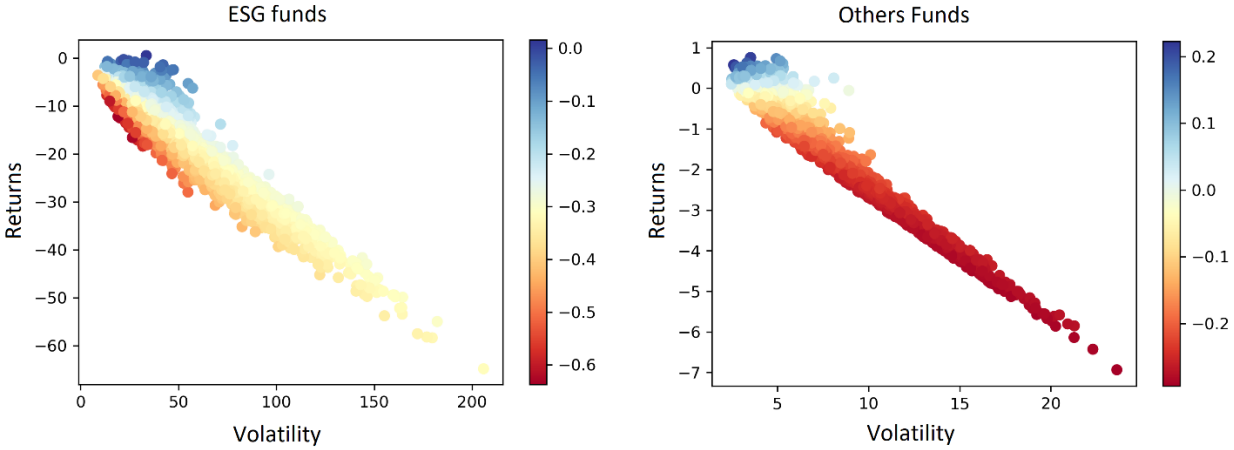


Figure 2 Portfolio Theory – Assessment of Evaluated Funds

The employed comparative approach in the form of a chi-squared test – good match test – was chosen to be a complement for reasons of verifying the correctness of the judged data. As we have mentioned, a suitable form might be the assessment of hundreds of input funds, not only tens. The applied approach also has a relatively good testifying value. The results of assessment of ESG and non-ESG funds is the contents of Table 3. From the values reached, the hypothesis of the independence of the selected rating on the evaluation of the performance in the chosen observed time-interval; i.e., we can assess the selected rating (method of evaluation) as satisfactory.

ESG Funds		6m	1yr	3yr	5yr
Rating	Pears. chí-sqr	2,916	5,000	2,916	5,000
	p	0,40	0,29	0,23	0,08

no-ESG Funds		6m	1yr	3yr	5yr
Rating	Pears. chí-sqr	5,000	5,000	1,875	5,000
	p	0,29	0,60	0,23	0,29

Table 3 Pearson's chi-squared applied on the funds data

4 Conclusions

The performed description is only a detailed sketch of the approach for the evaluation of the possible investment into funds, and this especially with SRI (ESG) evaluation, and a way of diversifying the risk of potential investments. A part of the solution of the project is also an effort to balance the convenience of the investments into the evaluated funds (SRI, ESG rating) vis á vis the unrated funds. The basis for any comparison is the availability of suitable descriptive data in the required form, to which, subsequently, suitable economic models are applied. Here, the approach of machine data-extraction from publicly available websites in the form of a created script (Morning Star, 2018) has proved to be competent. Thus, it is not necessary to have access to a paid database; this will be made use of as part of the created web-application designated to be a support platform for small investors who are considering investing into select funds.

The knowledge here acquired will be used in further research; collection of more input data and the creation of a greater input data-set will precede this research, so that it may be possible to apply this approach to the whole market-segment – namely, for instance, the Green Economy, Water Utilization, Sustainable Agriculture – in the form of ESG funds, thus enabling small investors to assess their potential investments in this sphere. Emphasis will be placed on the diversification of the investment risk.

Acknowledgements

The authors are thankful to the Grant Agency of Czech Republic (GAČR), project: "Modelling and simulation of sustainable investment decision-making", CEP ID GA17-23448S.

References

[1] Ameer, R. (2016). Usefulness of ASSET4 ESG for Socially Responsible Investing in a Small Economy. Journal of Investing. 25(2), pp 125–135.
[2] BNP Paribas, (2018). [cit. 2018-3-21] Available at: http://sustainable-responsible-investments.parvest.com/sri-home-en/.
[3] Chen, E., Nainggolan, Y. (2018). Distance bias of socially responsible investment. Social Responsinility Journal. 14(1). Pp 96–110. DOI: 10.1108/SRJ-02-2017-0021.
[4] Clark, G., Feiner, A., Viehs, M. (2015). From The Stockholder to The Stakeholder. University of Oxford. Oxford. p 63. [cit. 2018-3-21] Available at: https://arabesque.com/research/From_the_stockholder_to_the_stakeholder_web.pdf.
[5] Duuren, E., Plantinga, A., Scholtens, B. (2016). ESG Integration and the Investment Management Process: Fundamental Investing Reinvented. Journal of Business Ethic. 138(3), pp 525–533.
[6] Edwin, J. (2007). *Modern portfolio theory and investment analysis*. John Wiley & Sons, 2007. p 728. ISBN 0470050829.
[7] Erste group. (2018). SRRI (Synthetischer Risiko- und Ertragsindikator). [cit. 2018-3-21]. Available at: http://fondsabc.erste-am.com/srri/.
[8] Forbes. (2017). Socially-Responsible Investing: Earn Better Returns from Good Companies. [cit. 2018-3-21] Available at: https://www.forbes.com/sites/moneyshow/2017/08/16/socially-responsible-investing-earn-better-returns-from-good-companies/#5542a29623d0.
[9] Gera, I., Bahnhelyi, B. (2018). *Markowitz Portfolio Selection Using Various Estimators of Expected Returns and Filtering Techniques for Correlation Matrices*. Acta Polytechnica Hungarica. 15(1), pp 217–229.
[10] Geoffrey. (2013) *Five Questions about SRI – Weekly Expert Interview with Sonia Kowal, Director, Socially Responsible Investing, Zevin Asset Management,* [cit. 2018-3-21] Available at: http://www.emergingmarketsesg.net/esg/2013/08/02/five-questions-about-sri-weekly-expert-interview-with-sonia-kowal-director-socially-responsible-investing-zevin-asset-management-llc-boston-massachusetts-united-states-of-america/.
[11] Markowitz, H. (1954) *Portfolio Selection*. Journal of Finance. 1954. [cit. 2018-3-21] Available at: http://www.gacetafinanciera.com/TEORIARIESGO/MPS.pdf.
[12] Markowitz, H. (1959) *Portfolio Selection: Diversification of Investments*. John Wiley & Sons, Inc., New York. 1959. [cit. 2018-3-21] Available at: http://cowles.econ.yale.edu/P/cm/m16/index.htm.
[13] Meloun, M., Militký, J. Kompendium statistického zpracování dat. Praha: Karolinum, 2012. ISBN 978-80-246-2196-8.
[14] Morning Star (2018) [cit. 2018-3-21] Available at: http://www.morningstar.co.uk/uk/fundquickrank/default.aspx?category=EUCA000759.
[15] Python. 2017. Investment Portfolio Optimisation with Python. [cit. 2018-3-21] Available at: http://www.pythonforfinance.net/2017/01/21/investment-portfolio-optimisation-with-python/.
[16] Řezanková, H. Analýza dat z dotazníkových šetření. 3., aktualiz. vyd. Praha: Professional Publishing, 2011. ISBN 978-80-7431-062-1.
[17] Song, Y., Wang, H., Zhu, M. (2018). *Sustainable strategy for corporate governance based on the sentiment analysis of financial reports with CSR*. Financial Innovation. Springer. 4(1). DOI: 10.1186/s40854-018-0086-0.
[18] The Green New Deal. (2017). [cit. 2018-3-21] Available at: https://gpus.org/organizing-tools/the-green-new-deal/.
[19] Thomson Reuters. (2017). ESG research data. Thomson Reuters. 2018. [cit. 2018-3-21] Available at: https://financial.thomsonreuters.com/en/products/data-analytics/company-data/esg-research-data.html.
[20] USSIF – The Forum for Sustainable and Responsible Investment. (2018). Performance & SRI. [cit. 2018-3-21] Available at: https://www.ussif.org/performance.
[21] Voica, M., Panait, M., Radulescu, I. (2015). Green Investments - between necessity, fiscal constraints and profit. Bucharest. Procedia Economics and Finance. 22. pp. 72–79. DOI: 10.1016/S2212-5671(15)00228-2.

Behavior of the Czech small open economy in the conditions of occasionally binding constraints

Stanislav Tvrz[1], Jaromír Tonner[2], Osvald Vašíček[3]

Abstract. The period of economic recovery after the Great Recession was marked by utilization of non-standard monetary policy tools, such as quantitative easing or exchange rate interventions, by the central banks around the globe. In this paper we model the behavior of the Czech small open economy in the conditions of non-standard monetary policy with the use of DSGE model with occasionally binding constraints. We compare the baseline model to the alternative set-up, where the development of domestic and foreign policy interest rates and nominal exchange rate is restricted to given bounds. We investigate the behavior of model economy using the impulse response functions under the assumption that the limiting bounds are fully anticipated by the rational economic agents. We show that the effects of given constraints largely depend on the type of particular exogenous shock. While responses to some may remain unaffected, the rest gets to a certain extent amplified or dampened.

Keywords: DSGE model, simulation, non-standard monetary policy, occasionally binding constraints.

JEL classification: E32, E37
AMS classification: 91B51, 91B64

1 Introduction

The extraordinary conditions of the Great Recession of 2008-2009 brought about extraordinary measures. The standard monetary policy tools were exhausted as the policy interest rates hit near-zero values in reaction to the crisis around the globe. In the conditions of weak and inflation-less economic recovery, the central banks had to consider alternative ways of delivering further easing of monetary policy. Some, such as Federal Reserve System, European Central Bank or Bank of England launched quantitative easing programmes, while others, such as Swiss National Bank or the Czech National Bank introduced asymmetric commitments to keep their domestic currencies weaker than some given level of nominal exchange rate with respect to the euro.

The question of quantifying the estimated impact of non-standard monetary policy in this period for the Czech economy was studied e.g. in [2]. In this paper, we employ the method of modelling the occasionally binding constraints (OBC), as proposed in [4]. The same algorithm was applied before in [3], [7], [9] and [10]. Continuing this line of research, we focus on the behaviour of the model economy in this paper. We want to illustrate and better understand the implications of multiple simultaneous constraints, consistent with the situation during the CNB's exchange rate commitment period, for the response of the model economy to different exogenous shocks. Thus, we will introduce the zero lower bound in the domestic and foreign nominal interest rates and the asymmetric constraint in the nominal exchange rate of the Czech currency.

The rest of the paper is organised as follows: Section 2 describes the basic features of the DSGE model structure that is used in this paper; Section 3 introduces the very basics of the method of occasionally binding constraints simulation; Section 4 presents the impulse response functions simulation results; and Section 5 concludes.

2 Model

In this paper, we use the DSGE model structure in line with [1] and [6]. The relatively simple model structure contains representative optimizing economic agents of households and firms. There are intermediate firms and two type of retail firms, one selling domestic and one imported foreign goods. Households provide labour to the domestic intermediate firms and receive wages and dividends in return. Households spend their income on consumption but they can also use domestic bonds to smooth their consumption in time. Households and retail firms operate in monopolistically competitive markets. Therefore, they have some market power over their wages and prices. At the same time, Calvo-type nominal rigidity is assumed in both.

Following [8], the modified uncovered interest parity equation is assumed in the model:

[1]Czech National Bank, Monetary Department, Na Příkopě 28, 115 03 Praha, Stanislav.Tvrz@cnb.cz
[2]Czech National Bank, Monetary Department, Na Příkopě 28, 115 03 Praha, Jaromir.Tonner@cnb.cz
[3]Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, osvald@econ.muni.cz

$$E\left(e_{t+1}\right) - e_t = r_t - r_t^* + \gamma\sigma^2\left(f_t^* + f_t^{CB}\right), \tag{1}$$

where the expected appreciation of the nominal exchange rate on the left side of the equation is related with the nominal interest rate differential of domestic and foreign interest rates together with the capital inflows generated by the FX market dealers (f_t^*) and the interventions of the domestic central bank (f_t^{CB}). Parameter γ stand for the risk aversion of the FX market dealers, while the parameter σ denotes the standard deviation of the nominal exchange rate. The capital inflows f_t^* are given exogenously and are modelled as an AR(1) process. Interventions of the central bank are modelled in line with the asymmetric commitment of the CNB to keep the nominal exchange rate at a level not stronger that 27 CZK/EUR, which means that they contain an occasionally binding constraint:

$$f_t^{CB} = \min\left\{0, \chi^e(e_t - \widetilde{e}) + \varepsilon_t^{f_{CB}}\right\}, \tag{2}$$

where χ^e is the elasticity of the central bank's reaction to the deviations of the nominal exchange rate from its target $\widetilde{e}$ and $\varepsilon_t^{f_{CB}}$ is exogenous shock in central bank's interventions.

Monetary policy of the central bank is represented by a Taylor rule with zero lower bound:

$$r_t = \max\left\{0, \rho^{MP} r_{t-1} + \left(1-\rho^{MP}\right)\left(\phi^\pi E(\pi_{t+1}) + \phi^p p_t + \phi^y y_t + \phi^e e_t\right) + \varepsilon_t^{MP}\right\}. \tag{3}$$

Thus, the central bank can in general react to deviations in expected inflation $E(\pi_{t+1})$, price level p_t, output y_t or nominal exchange rate e_t, depending on values of specific elasticity parameters $\phi^x, x \in \{\pi, p, y, e\}$. ρ^{MP} stands for interest rate smoothing parameter and ε_t^{MP} is exogenous i.i.d. monetary policy shock. Price level targeting is included in line with [5] in order to make the OBC algorithm more stable.

Foreign sector is modelled in a structural way. It contains analogous equations to domestic New Keynesian Phillips curve, dynamic IS curve, marginal costs and the Taylor rule equations. A zero lower bound is assumed in the foreign sector in the same way as in (3).

3 Method

Following [4] we estimate the model with occasionally binding constraints. The algorithm generates a sequence of shadow price shocks that allows the bounded variables to remain in specified bounds. In an otherwise linear model, a bounded variable $x_{1,t}$ can be specified e.g. as

$$x_{1,t} = \max\left\{0, \mu_1 + \phi_{-1}x_{t-1} + \phi_0 x_t + \phi_1 E(x_{t+1}) - (\phi_{-1} + \phi_0 + \phi_1)\mu\right\}, \tag{4}$$

where x_t is a vector of model variables, μ is a vector of their steady states, and $\mu_1, \phi_{-1}, \phi_0, \phi_1$ are vectors, $\mu_1 > 0$. Supposing that the length of impulse response vector is T and the last period where the constraint is binding is $T^*, T^* \leq T$, the algorithm finds a sequence of shadow price shocks $\epsilon_{s,t-s}^{SP}$, s.t. following equation fulfils the constraint (4):

$$x_{1,t} = \mu_1 + \phi_{-1}x_{t-1} + \phi_0 x_t + \phi_1 E(x_{t+1}) - (\phi_{-1} + \phi_0 + \phi_1)\mu + \sum_{s=0}^{T^*-1} \epsilon_{s,t-s}^{SP}. \tag{5}$$

The shadow price shocks ϵ_s^{SP}, t act as news shocks - they are known at time t, but hit the economy in $t+s$. In the impulse response exercise, only the values $\varepsilon_{s,0}^{SP}$ are different from zero, because all the news arrive in period $t = 0$.

4 Results

In this section we get to the main contribution of the paper, i.e. the economic interpretation of the model behaviour as captured by the impulse responses to different exogenous structural shocks under the simultaneous occasionally binding constraints on nominal exchange rate e_t, domestic nominal interest rate r_t and foreign nominal interest rate r_t^*. IRFs are presented in Figures 1 through 7 and they are depicted as logarithmic deviations from steady state.

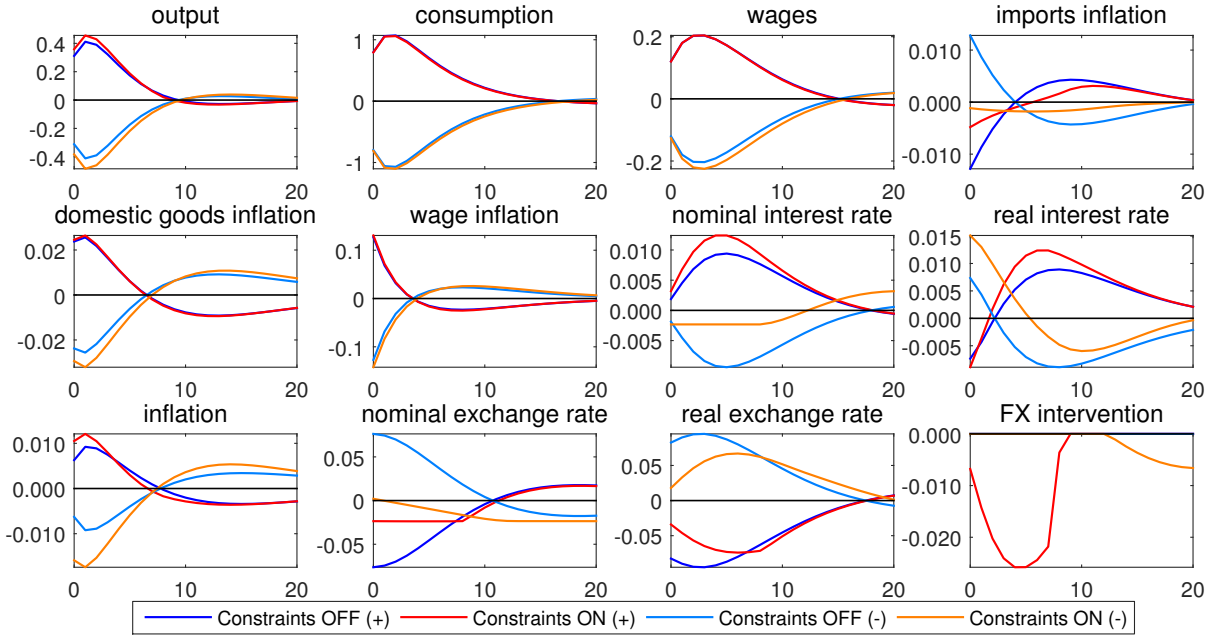


Figure 1 Demand shock

Occasionally binding constraints strengthen the effects of a demand shock (Figure 1), i.e. they work in a pro-cyclical way. When the shock is positive, the exchange rate appreciation is constrained in its reaction to higher interest rates and the output gap is more positive compared to the situation without any constraint. In the case of a negative demand shock, the monetary conditions cannot be relaxed sufficiently via lower interest rates, thus the output gap is more negative compared to the situation without constraints.

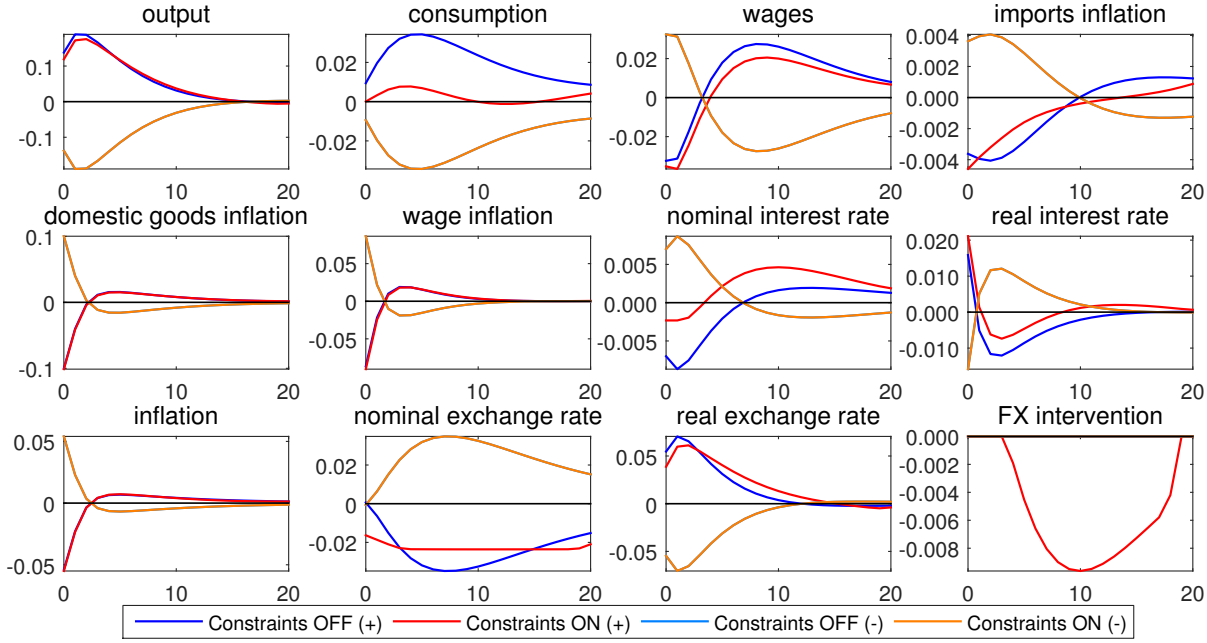


Figure 2 Technology shock

In the case of a technology shock (Figure 2), occasionally binding constraints work against the cycle. When the shock is positive, cutting of interest rates is limited, thus the output gap does not open as much. When the shock is negative, the monetary conditions can be relaxed sufficiently by exchange rate depreciation along with raising interest rates.

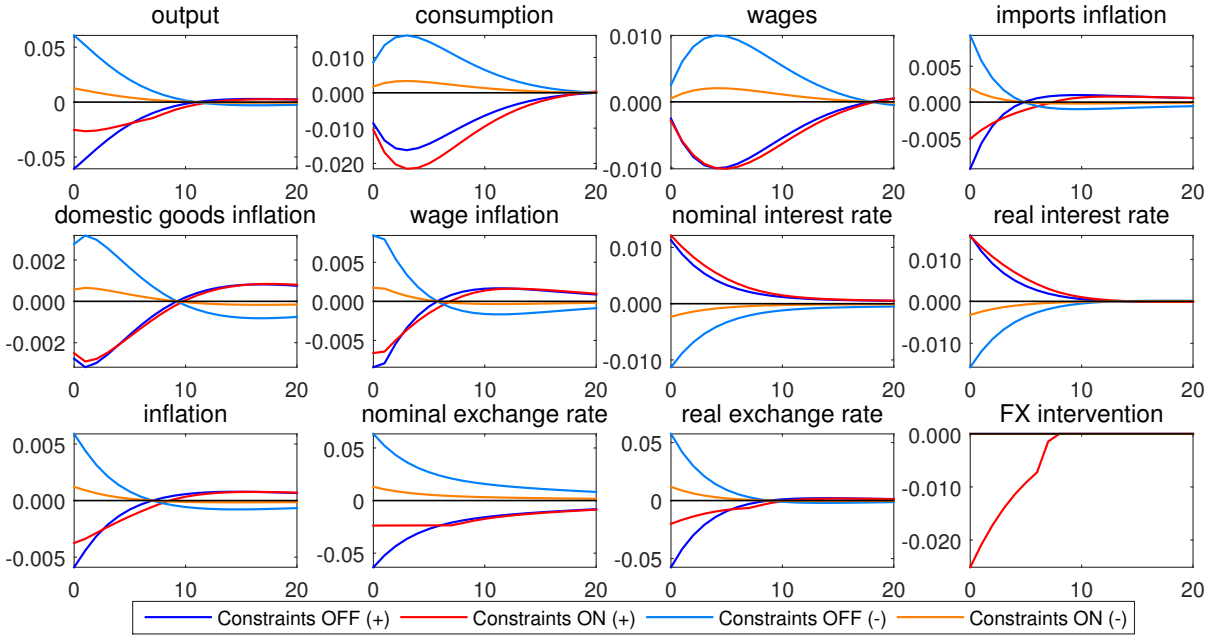


Figure 3 Monetary policy shock

When a monetary policy shock is considered (Figure 3), occasionally binding constraints work against the cycle. A positive shock does not lower the output gap as much, because the exchange rate cannot appreciate sufficiently. A negative shock hits the zero lower bound, thus exchange rate cannot depreciate as much and output gap is little less positive.

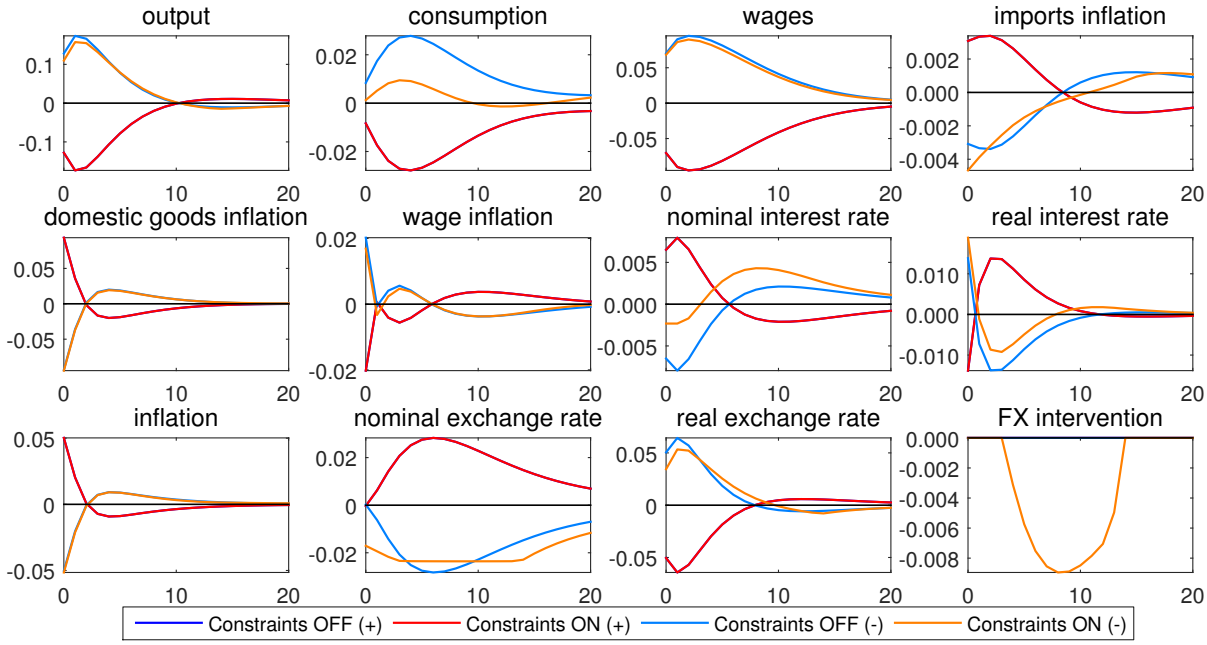


Figure 4 Cost push shock

For a cost push shock (Figure 4), constraints work in an asymmetric way. When a positive shock hits the economy, constraints do not apply because exchange rate depreciates and interest rate increases. For a negative shock, both constraints apply and output gap increases less because exchange rate appreciates less compared to the case without constraints.

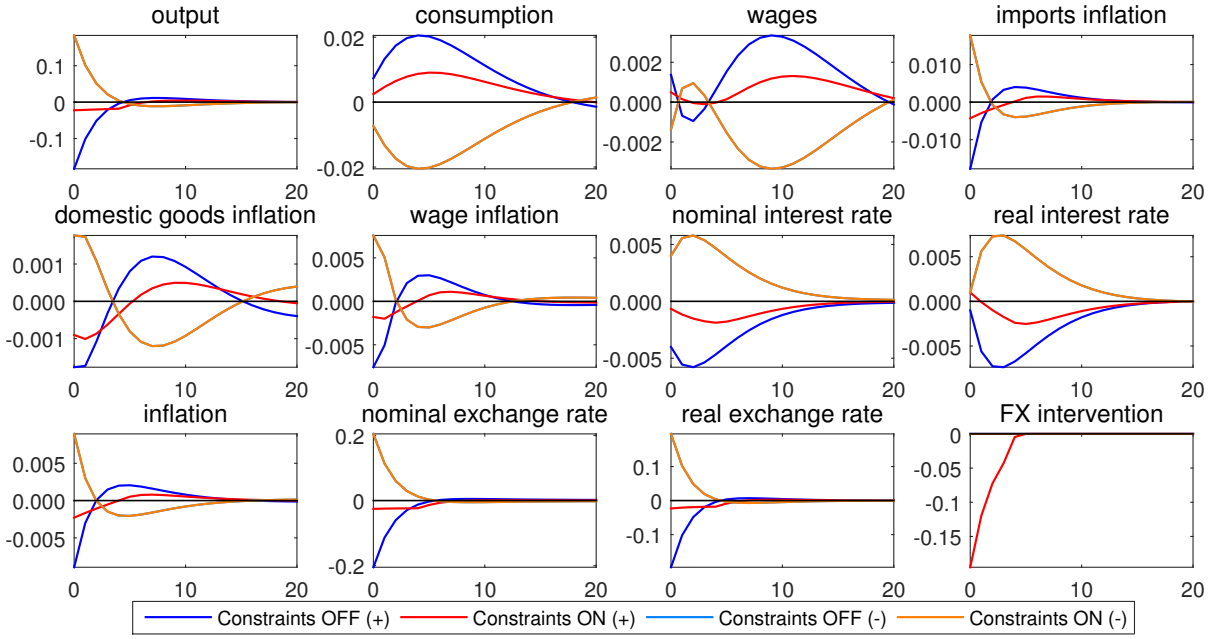


Figure 5 Capital flows shock

An outflow of capital (IRF in Figure 5) depreciates the exchange rate and increases interest rates, thus constraints do not apply. In the case of capital inflow, its effects are counter-cyclical as interest rates cannot decrease as much.

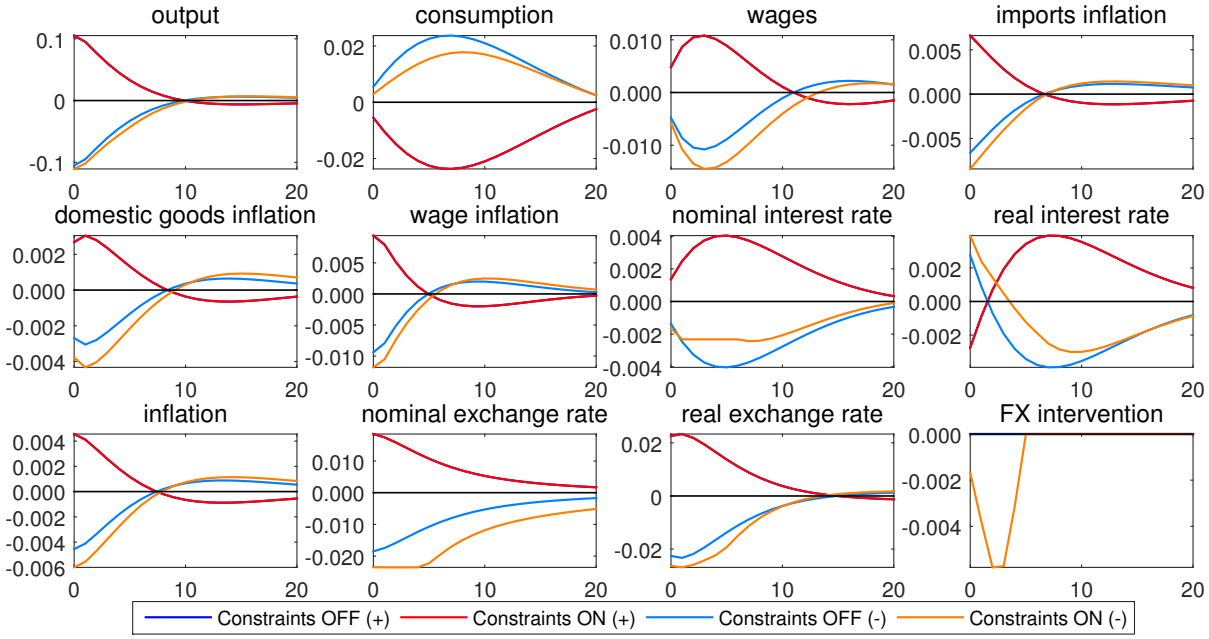


Figure 6 Foreign demand shock

A positive foreign demand shock (Figure 6) increases interest rates and makes exchange rate weaker, thus constraints do not apply. For a negative shock, their effects are pro-cyclical due to higher interest rates and stronger exchange rate.

A positive foreign monetary policy shock (Figure 7) does not collide with the constraints, because it increases the interest rate and makes the exchange rate weaker. The effect on the output gap is a puzzle. For a negative shock, the effects of the constraints are counter-cyclical.

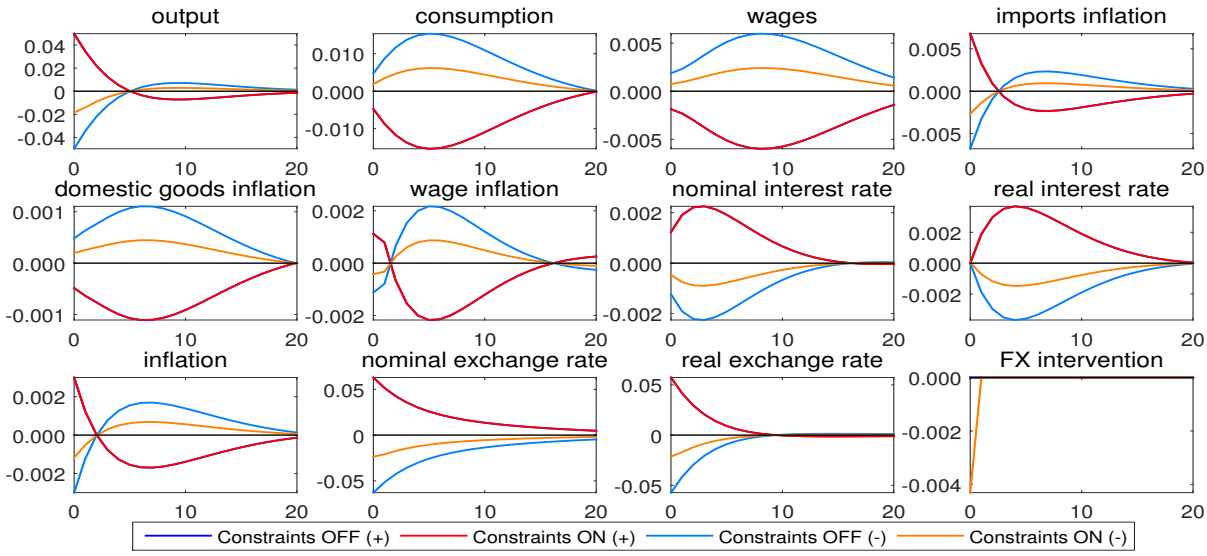


Figure 7 Foreign monetary policy shock

5 Conclusion

We have shown the possible implications of multiple occasionally binding constraints in the domestic and foreign economy at the same time. The main result of this paper is that the results depend largely on the type of particular exogenous shock in question. While some shocks may remain unaffected, in other cases, the constraints may dampen or amplify the response of the domestic economy in comparison to the baseline model.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/0966/2017. This support is gratefully acknowledged. The views expressed here are those of the authors and do not necessarily reflect the position of the Czech National Bank.

References

[1] Alpanda, S., Kotzé, K. & Woglom, G. (2010). The role of the exchange rate in a New Keynesian DSGE model for the South African economy. *South African Journal of Economics*, 78, 170-191.

[2] Brůha, J. & Tonner, J. (2017). *An Exchange Rate Floor as an Instrument of Monetary Policy: An Ex-post Assessment of the Czech Experience*. CNB Working Paper Series, 4/2017.

[3] Fodor, J. & Vašíček, O. (2017). Unconventional monetary policy in a small open economy under inflation targeting regime. In P. Pražák (Ed.), *35th International Conference Mathematical Methods in Economics MME 2017 Conference Proceedings* (pp. 177-182). Hradec Králové: University of Hradec Králové.

[4] Holden, T. & Paetz, M. (2012). *Efficient simulation of DSGE models with inequality constraints*. Guildford: University of Surrey. Discussion Papers in Economics 16/2012.

[5] Holden, T. (2017). *Existence, uniqueness and computation of solutions to dynamic models with occasionally binding constraints.* Guildford: University of Surrey. Preprint.

[6] Justiniano, A. & Preston, B. (2010). *Monetary policy and uncertainty in an empirical small open economy model.* Working Paper Series WP-09-21, Federal Reserve Bank of Chicago.

[7] Malovaná, S. (2015). *Foreign Exchange Interventions at the Zero Lower Bound in the Czech Economy: A DSGE Approach.* Prague: Charles University. Working Papers IES 2015/13.

[8] Montoro, C. & Ortiz, M. (2013). *Foreign Exchange Intervention and Monetary Policy Design: A Market Microstructure Analysis.* Basel: Bank for International Settlements. Draft.

[9] Súkupová, K. & Vašíček, O. (2016). Modelling of an asymmetric foreign exchange rate commitment in the Czech economy: DSGE model with constraints. In A. Kocourek & M. Vavroušek (Eds.), *34th International Conference Mathematical Methods in Economics 2016 Conference Proceedings* (pp. 797–802). Liberec: Technical University of Liberec.

[10] Súkupová, K. & Vašíček, O. (2016). Multiple constraints applied into the DSGE model: The case of unconventional monetary easing in the Czech economy. In M. Reiff & P. Gežík (Eds.), *Proceedings of the International Scientific Conference Quantitative Methods in Economics Multiple Criteria Decision Making XVIII*. (pp. 356–361). Bratislava: University of Economics.

Remarks on De Novo approach to multiple criteria optimization

Milan Vlach[1], Helena Brozova[2]

[1]*Institute of Information Theory and Automation, Pod Vodarenskou vezi 4, Prague, and Charles University, Faculty of Mathematics and Physics, Department of Theoretical Computer Science and Mathematical Logic,*
vlach@utia.cas.cz or milan.vlach@mff.cuni.cz
[2]*Czech University of Life Sciences, Kamycka 129, Prague, Czech Republic,*
brozova@pef.czu.cz

Abstract. During the 1980's and 1990's Milan Zeleny, in a number of papers, proposed and developed a method (called De Novo Programming) for embedding a given linear optimization model into a certain class of linear models by allowing for reshaping the feasible set of the original problem. The method uses, in an essential way, a transformation of the original problem to continuous knapsack problem, and it concerns only models with capacity constraints and some implicit assumptions about the problem data.
First we briefly recall De Novo methodology and present conditions on the initial data – values of cost coefficients, technical coefficients, prices of resources and total available budget - that are necessary for applicability. Then we present an example of adaptation of De Novo approach for models with both capacity and requirement constraints. In such cases the transformation to continuous knapsack problem is not possible. But the so called optimum-path ratio for achieving the best performance for a given budget B is again applicable.

Keywords: De Novo programming, model conditions, Pareto solution, capacity and requirement optimization.

JEL Classification: C61
AMS Classification: 90C29

1 Introduction

1.1 De Novo approach to single criterion optimization

To motivate the De Novo approach ([5], [6], [7], [8]) to modeling and analyzing decision making, let us first consider one of the standard linear programming models for allocating resources to activities so as to attain a given economic objective; for example

$$c^T x \to Max \quad \text{subject to} \quad Ax \leq b, x \geq 0 \tag{1}$$

where A is a real (m, n)-matrix, b is a real m-vector, and c is a real n-vector.

Since the early days of linear programming, both practitioners and theorists have been interested in behaviour of solutions when coefficients of the problem are allowed to vary. Such questions have led to the emergence of sensitivity analysis (investigation of changes in the individual coefficients which cause an optimal solution to become nonoptimal) and parametric programming (investigation of changes when some of the coefficients are assumed to be linear functions of a parameter).

The De Novo methodology ([6], [7]) also deals with changes in some of the input data; in particular, it deals with the changes in the components b_i of the right hand side vector b of model (1). Clearly, such changes modify the set of feasible solutions, which may change the set of maximizers. In contrast to the sensitivity analysis and parametric programming, De Novo requires additional exogenous data, cost of resources and level of available budget. If p denotes a given m-vector of unit cost of resources and B denotes a given available budget, then De Novo approach gives to b the freedom to vary freely in the region given by

$$pb \leq B \text{ and } b \geq 0. \tag{2}$$

To indicate that the components of b are now real variables we change the notation and use the letter y instead of b. Therefore, instead considering problem (1) with a fixed b, we deal with the problem

$$c^T x \to Max \quad \text{subject to} \quad Ax - y \leq 0, py \leq B, x \geq 0, y \geq 0. \tag{3}$$

This is again a linear programming problem, which can be solved by any procedure for solving linear programming problems. To refer to the special structure of this problem, we will say that we are dealing with *De Novo programming problem with a single criterion*. The De Novo approach takes advantage of the fact that (under certain assumptions about A, B, c and p) problem (3) can be rewritten as follows:

$$c^T x \to Max \quad \text{subject to} \quad pAx \leq py, py \leq B, x \geq 0, y \geq 0. \tag{4}$$

To solve this problem the De Novo approach uses an obvious observation that for each feasible solution (x,y) of problem (4), x is also a feasible solution of the problem

$$c^T x \to Max \quad \text{subject to} \quad Vx \leq B, x \geq 0 \tag{5}$$

where V stands for the n-vector pA.

Similarly to problem (3), problems (4) and (5) are specially structured linear programming problems that can be solved and whose properties can be studied by standard linear programming means. However, it is useful to take advantage of their special structure. In particular, the De Novo approach takes advantage of the fact that problem (5) is a continuous linear knapsack problem whose optimal solution $\hat{x}$ can easily be obtained (for strictly positive components of c and of V) as follows: Let c_k/v_k be such that $c_k/v_k = max\{c_1/v_1, c_2/v_2, \dots, c_n/v_n\}$. Then the components of the optimal solution $\hat{x}$ are given by $\hat{x}_i = B/v_i$, when $i = k$, and $\hat{x}_i = 0$ otherwise. Using $\hat{x}$, we set $\hat{y} = A\hat{x}$ and $\hat{B} = V\hat{x}$. The resulting triple $(\hat{x}, \hat{y}, \hat{B})$ is then called the *optimal system design* for De Novo problem (3). Note that, in our simple situation with single objective function, we have $\hat{B} = B$.

1.2 De Novo approach to multiple criteria optimization

In practice, many (if not most) of decision problems have multiple objectives to be optimized simultaneously, and the De Novo approach has been successfully used for designing and solving a number of such problems ([1], [2], [3], [4], [5], [6], [7], [8]). In such situations problem (1) is replaced by the problem

$$Cx \to Max \quad \text{subject to} \quad Ax \leq b, x \geq 0 \tag{6}$$

where C is a real (q, n)-matrix of coefficients of q objective functions. Now the situation is more complicated because one has to take into account the fact that the individual objective functions can attain their optimal values on the common feasible set in different points.

To take care of this complication, the De Novo approach first proceeds as in the single criteria problems (replacing c by C) until solving the corresponding knapsack problems. Let Z^*be the q-vector of optimal values of individual objective functions over the set of feasible solutions, which is the same as in problem (5). Then Z^* is used to define an auxiliary problem, called the *meta-optimization problem*, which is formulated as the minimization problem

$$Vx \to Min \quad \text{subject to} \quad Cx \geq Z^*, x \geq 0. \tag{7}$$

Let x^* be the resulting minimizer, and let define y^* and B^* by $y^* = Ax^*$and $B^* = Vx^*$. It is easy to see that $B^* \geq B$ and that the value B^* is the minimum budget for obtaining at least Z^* by using x^* and y^*.
The fraction $r^* = B/B^*$ is called *the optimum path ratio* ([4], [6], [7]) for approaching $Z^*$with respect to given budget B, and the ordered triple (r^*x^*, r^*y^*, r^*Z^*) is called the *optimal system design* for the following De Novo problem

$$Cx \to Max \quad \text{subject to} \quad Ax - y \leq 0, py \leq B, x \geq 0, y \geq 0. \tag{8}$$

The aim of our paper is to show how and under what conditions on the initial data (values of objective functions coefficients, technical coefficients, prices of resources, and total available budget) the De Novo programming methodology is functioning in problems with multiple criteria. We also show when the approach can fail, and that not always it is possible to solve the model (8) with the help of knapsack problem. Furthermore we present the adaptation of De Novo approach to models with both capacity and requirement constraints, and conclude with some comments about possible future developments.

2 Conditions for De Novo approach

2.1 Remarks on the De Novo optimization application

Remark 1

De Novo optimization is proposed as a tool for optimization of system design. Therefore, for reasonability of the decision-making process, it is natural to assume that $B = pb$ for given initial model (1), given budget B, and given price vector p of resources. As mentioned previously, the De Novo approach transforms problem (3) into problem

(4). This is possible, because every feasible solution (x, y) of problem (3) is also a feasible solution of problem (4) and the coefficients at the components of y are zero. It is also not difficult to see that if (x, y) is a feasible solution of problem (4), then x is a feasible solution of the problem (5). Moreover, since to each feasible solution of problem (5), it corresponds at least one (usually many) y such that (x, y) is a feasible solution of problem (4) and because all coefficients at the components of y are zero, it is possible (under certain conditions, see the next remark) to deal with problem (5).

Remark 2

Implicit assumption, reasonable for a large class of problems, is that B and all components of p are positive. However, almost none of the published articles mentions that the transformation of the model (4) to the knapsack problem (5) requires the positivity of components of pA because problem (5) may have no solution if some components of pA are negative or zero. Notice that the positivity of pA is guaranteed if matrix A is nonnegative and has no zero-column and all components of vector p are positive.

As a simple illustration, let us consider the following example of optimization of system design, which has to be solved under the condition of the total budget $B = 5$ and, prices of resources equal to $p = (1, 1, 1)$ and the initial problem

$$x_1 \to Max \text{ subject to } \begin{pmatrix} -1 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \leq \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, x \geq 0.$$

Using De Novo approach, we receive $pA = (0, 0, 1)$ and the initial model will be reformulated into the following instance of continuous knapsack problem (5)

$$1x_1 \to Max \text{ subject to } (0 \quad 0 \quad 1) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \leq 5, x \geq 0.$$

Now the objective function of this knapsack problem is unbounded on the feasible set, and therefore no maximal value of it exists. Nevertheless, if only the transformation of problem (3) to problem (4) would be used, then the solution would exist. Indeed, the problem (4) takes the form

$$x_1 \to Max \text{ subject to } \begin{pmatrix} -1 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \; y_1 + y_1 + y_1 \leq 5, \; x, y \geq 0$$

and has the solution $x = (5, 0, 0)$ and $b = (0, 5, 0)$. The optimal system design under the budget $B = 5$ is given by the problem

$$x_1 \to Max \text{ subject to } \begin{pmatrix} -1 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \leq \begin{pmatrix} 0 \\ 5 \\ 0 \end{pmatrix}, x \geq 0.$$

Remark 3

The De Novo model construction suppose only the constraints of type $Ax \leq b$, so called the capacity constraints, and not the constraints of type $Ax \geq b$, so called the requirements constraints. The requirements constraints for example ensure the required amount of production for sale. In practical applications, it is often necessary to assume the cost of such requirements (the cost of the contract signed). If the De Novo standard procedure is used, the optimum value of the requirements will be zero, which corresponds to the general opinion that the requirements have to be neither set nor even paid. In practice, it can lead to an optimal system design not having contracts for production, but very often, such contracts are necessary. A way to overcome this problem is given in the next chapter.

2.2 De Novo approach for model with requirement constraints

De Novo approach allows to decision-maker to optimize not only values of criteria function but also to optimize the system design under given budget. Zeleny ([5], [6], [7], [8]) implements this approach for models with capacity constraints only. Because it is often necessary to optimize the system described not only by capacity constraints but also by requirement constraint, we propose the following generalization of De Novo optimization approach. In this case, the optimal values of capacities and demands under given budget will be found.
The typical multiple criteria linear optimization model with the capacity and requirement constraints can be written as follows

$$Cx \to Max \quad \text{subject to} \quad A^{\leq}x \leq b^{\leq}, \; A^{\geq}x \geq b^{\geq}, \; x \geq 0 \tag{9}$$

where $A^{\leq}, b^{\leq}$ are coefficients from capacity constraints
$A^{\geq}, b^{\geq}$ are coefficients from requirement constraints and
C are coefficients of cost functions.

The criteria optimization means to find the optimal values of criteria functions. The system design optimization means to find the optimal values of capacities and requirements under the given budget.
Consider now the values of capacities and requirements as variables and reformulate constraints as follows:

- capacity constraints with unknown capacities

$$A^{\leq}x - y^{\leq} \leq 0 \tag{10}$$

- requirements constraints with unknown requirements

$$A^{\geq}x - y^{\geq} \geq 0 \tag{11}$$

- cost of necessary capacities and possible requirements has to be less or equal to the given budget B

$$p^{\leq}y^{\leq} + p^{\geq}y^{\geq} \leq B \tag{12}$$

where $y = (y^{\leq}, y^{\geq})$ are unknown values of capacities and requirements, $p^{\leq}, p^{\geq}$ are the cost of capacities and requirements and B is the available budget.

Optimal budget allocation means looking for optimal necessary capacities and possible requirements. If these values are known, the constraints of type (10) and (11) can be seen as the equations ([8]). Into the set of constrains, the budget constraint (12) has to be added. New model formulation will be

$$Cx \to Max \quad \text{subject to} \quad A^{\leq}x - y^{\leq} = 0\,,\ A^{\geq}x - y^{\geq} = 0,\ p^{\leq}y^{\leq} + p^{\geq}y^{\geq} \leq B,\ x, y \geq 0 \tag{13}$$

The partial optimal solutions of model (13) are found for each criterion and the ideal values of all criteria create the vector

$$Z^I = (z_1^I, \dots, z_q^I) \tag{14}$$

The new problem now is to find the minimal necessary budget guarantee at least ideal values of criteria, the formulation of this meta-optimum model is

$$p^{\leq}y^{\leq} + p^{\geq}y^{\geq} \to Min \quad \text{subject to} \quad A^{\leq}x - y^{\leq} = 0\,,\ A^{\geq}x - y^{\geq} = 0,\ Cx \geq Z^I,\ x, y \geq 0 \tag{15}$$

After solving the model (15), the minimal budget for achieving ideal criteria values B^*, the optimal solution (x^*, b^*) and achieved values of criterial functions $Z^* = Cx^*$ are received. Generally this minimal budget B^* is greater than available budget B. The optimum-path ratio for achieving the best performance Z for a given budget B can be defined using for instance ratio $r^* = \frac{B}{B^*}$. Other possible optimum-path ratio formulas can be found in the paper of Shi ([4]). By using optimum-path ratio r^*, the following data for optimal system design can be received:

Optimal right hand side values $b = r^*b^*$

Optimal values of variables $x = r^*x^*$

Optimal values of criteria $Z = r^*Cx^*$

The optimal system design $A^{\leq}x = (r^*b^*)^{\leq}, A^{\geq}x = (r^*b^*)^{\geq}, x \geq 0$

2.3 Example

The above-mentioned generalization of De Novo optimization is now illustrated in the following example:

$$\begin{pmatrix}1 & 1\\ 1 & 4\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} \to Max \quad \text{subject to} \quad \begin{pmatrix}3 & 4\\ 1 & 3\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} \leq \begin{pmatrix}60\\ 30\end{pmatrix}$$
$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} \geq \begin{pmatrix}5\\ 5\end{pmatrix}$$
$$x \geq 0$$

The set of feasible solutions and the partial optimal solutions for both criteria under fixed system design are in the Figure 1.

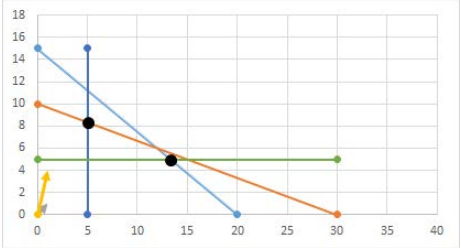


Figure 1 Set of feasible solutions and partial optimal solutions

Let the available budget be $B = 43$ and the costs of capacities and requirements be $p = (0.5, 0.4, 0.1, 0.1)$. The reformulation of the initial model for the optimal budget allocation is following

$$\begin{pmatrix}1 & 1\\ 1 & 4\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} \to Max \quad \text{subject to} \quad \begin{pmatrix}3 & 4\\ 1 & 3\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} - \begin{pmatrix}y_1\\ y_2\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}$$
$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} - \begin{pmatrix}y_3\\ y_4\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}$$
$$0.5y_1 + 0.4y_2 + 0.1y_3 + 0.1y_4 \leq 43$$
$$x, y \geq 0$$

The ideal values of both criteria create the ideal vector $Z^I = (21.5, 52.12)$ and the partial optimal solutions for both criteria under the available budget are in the Figures 2 and 3.

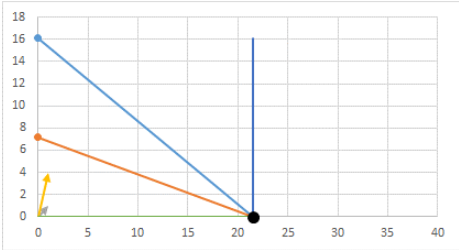


Figure 2 Optimal system design and optimal solution for the first criterion

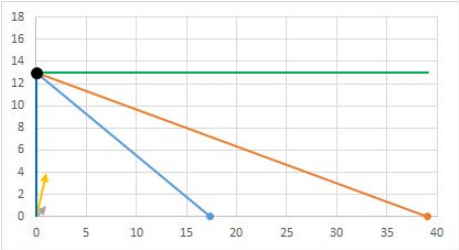


Figure 3 Optimal system design and optimal solution for the second criterion

The minimal necessary budget for achievement of ideal criterial values is computed using the following model

$$0.5y_1 + 0.4y_2 + 0.1y_3 + 0.1y_4 \to Min \quad \text{subject to} \quad \begin{pmatrix}3 & 4\\ 1 & 3\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} - \begin{pmatrix}y_1\\ y_2\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}$$
$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} - \begin{pmatrix}y_3\\ y_4\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}$$
$$\begin{pmatrix}1 & 1\\ 1 & 4\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} \geq \begin{pmatrix}21.5\\ 52.12\end{pmatrix}$$
$$x, y \geq 0$$

Optimal system design needs budget $B^* = 56.269$, optimal solution is $x^* = (11.293, 10.207)$, and right hand sides are $b^* = (74.707, 41.914, 11.293, 10.207)$. Because the necessary budget is higher than available, the optimum-path ratio is $r = \frac{43}{56.269}$. Therefore the achievable optimal solution and optimal system design is described as $x^* = (6.63, 7.8)$, and $b^* = (57.09, 32.03, 8.63, 7.8)$. Optimal values of criteria are $Z^{I*} = (16.43, 39.83)$. See Figure 4 and 5.

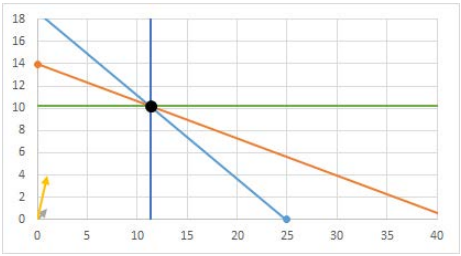


Figure 4 Optimal system design with minimal budget for ideal criteria values

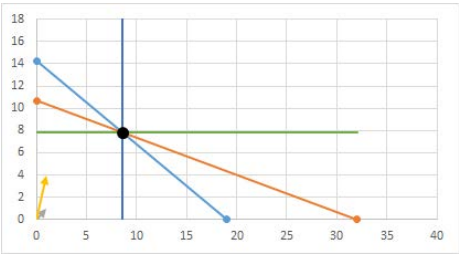


Figure 5 De Novo system design with maximal possible criterion values under given budget

The criterial space of the solved problems is in the Figure 6. The blue line bounds the set of possible criterial values under the initial constraints and the grey dotted line describes the partial optimal values of criteria. The orange line shows possible criterial values of both criteria achievable with different budget and with optimal system design. It is seen, that optimization of system design allows to reach higher (or better) criteria values than that under the initial fixed constraints.

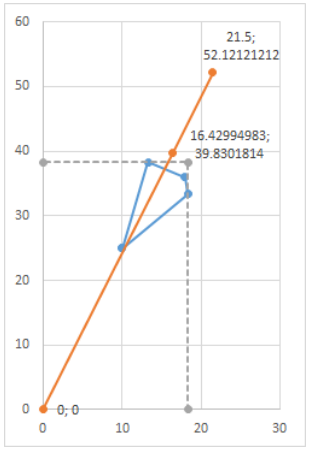


Figure 6 Criterial space and solutions of model

3 Conclusion

We have seen, the De Novo approach uses several problem transformation to obtain the optimum path ratio and the corresponding optimal design for the De Novo problem with single or multiple criteria. All of these transformations require that the values of individual components of y can vary freely in the region given by (2). This is certainly not always possible in practice, and the validity of this assumption may be lost with increasing value of parameter B. Thus it would be valuable to investigate possibilities for including some reasonable restrictions of the region given by (2).

It seems that the transformation from (3) to (4) may be related to duality of linear programming, which could be of interest both from the theoretical and practical point of view.

We have also seen that it will be of interest to study the classes of problems that cannot be solved by transformation to the knapsack problem and develop for them specialized procedures.

There are also some alternative proposals for introducing the optimum path ratios ([4]) and therefore for optimal system designs. It seems that the terminology is slightly confusing: Which optimal one would be optimal? It would be desirable to clarify relations between different notions of optimum path ratios and optimal system design, and search for others.

Acknowledgements

This research is supported by the grant of the Czech Science Foundation No. P402/12/G097 ‘DYME - Dynamic Models in Economics’.

References

[1] Babic, Z., Pavic, I. (1996). Multicriterial production planning by De Novo programming approach. *International Journal of Production Economics*, 43(1), 59–66.
[2] Fiala, P. (2011). Multiobjective De Novo Linear Programming. *Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica*, 50(2), 29–36.
[3] Huang, Ch.Y., Tzeng, G.H. (2007). Post-Merger High Technology R&D Human Resources Optimization Through the De Novo Perspective. In: *Advances in Multiple Criteria Decision Making and Human Systems Management: Knowledge and Wisdom*, Ed. Y. Shi et al., IOS Press, 47-64.
[4] Shi, Y. (1995). Studies on Optimum-Path Ratios in De Novo Programming Problems. *Computers and Mathematics with Applications*, 29(5), 43–50.
[5] Zeleny, M. (2010). Multiobjective Optimization, Systems Design and De Novo Programming. In: Zopounidis C., Pardalos P. M. (eds.): *Handbook of Multicriteria Analysis*, Berlin: Springer.
[6] Zeleny, M. (1990a). De Novo Programming. *Ekonomicko-matematický obzor*, 26(4), 406–413.
[7] Zeleny, M. (1990b). Optimal Given System vs. Designing Optimal System: The De Novo Programming Approach. *International Journal of General System*, 17(4), 295–307.
[8] Zeleny, M. (1986). Optimal system design with multiple criteria: De Novo programming approach. *Engineering Costs and Production Economics*, 10(2), 89-94, 1986.

Problem of competing risks with covariates: Application to an unemployment study

Petr Volf[1]

Abstract. This study deals with the methods of statistical analysis in the situation of competing risks in the presence of regression. First, the problem of identification of marginal and joint distributions of competing random variables is recalled. The main objective is then to demonstrate that the parameters and, in particular, the correlation of competing variables, may depend on covariates. The approach is applied to solution of a real example with unemployment data. The model uses the Gauss copula and Cox's regression model.

Keywords: statistical survival analysis, competing risks, copula, unemployment study, Cox's regression model.

JEL classification: C41, J64
AMS classification: 62N02, 62P25

1 Introduction and Motivation

The problem of competing risks arises when two (or more) mutually dependent random times to certain events (e.g. a failure of a device which can be caused by two different reasons) are followed and just the first (least) of them is registered. The phenomenon occurs frequently in areas of reliability, biostatistics and medical studies, as well as in demography studies, labor statistics, insurance, and in econometrics generally. The interest in the problem dates back to 70-ties of the last century. It is well known that, in general, without additional assumptions (e.g. on a parametric form of distributions) the model is not identifiable. It means that for each level of dependence we may obtain different estimate of joint probability distribution of both random times, and we are not able to choose among them. Further, Heckman and Honoré (1989) have shown (rather than proved, as their argumentation is more verbal than precisely mathematical) that when an additional information on covariates is available, identification is possible. Namely, they dealt with competing times with Cox's and AFT regression models. Their result is in the background of the unemployment duration study by Han and Hausman (1990), with two competing events of unemployment termination. However, they quite neglected the possibility that the parameter characterizing mutual dependence of risks (e.g. the correlation) may also depend on the covariate. When this is the case, the problem of non-identifiability arises anew. Later on the case of competing risks with covariates was solved by many other authors, in a more precise way, already with the aid of a copula describing the dependence, see e.g. Lee (2006), Berg et al. (2007). However, the problem of possible dependence of copula parameter on the covariates was not discussed. That is why the present paper opens this problem. On another unemployment data, which are commonly available on the Web (as the original data used in Han and Hausman are not available to me), the model of joint distribution is formulated, with the use of Gauss copula. Then, it is shown that the correlation, which characterizes the dependence of two random times, depends on the age of employee, here taken as a covariate. In fact, the result of Heckman and Honoré and of others can be used for at least approximate estimation of partially constant correlation in a moving window scheme. In such a way its dependence on the covariate can be demonstrated.

The outline of the paper is the following: The next section introduces the scheme of competing risks, presents the method of analysis of competing events incidence, and points to the problem of possible non-identifiability of their marginal probability distributions. We shall mention also certain identifiability results in the framework of regression models. Then the notion of copula is recalled and used in competing risks model formulation. In Section 3 the Gauss copula is introduced and the procedure of simultaneous maximum likelihood estimation of marginal distributions and correlation is described. Finally, in Section 4, the Gauss copula and Cox's regression model are jointly applied to a real example. We use the data taken from Kadane and Woodworth (2004) recording the employment and its termination in certain company. There are two competing events, dismissal or voluntary leaving the job, it is expected that their risks are dependent mutually and also on the age of employees, which is taken as a covariate. The objective is to show that the model parameters and, in particular, the correlation of both risks, depend on this covariate. The solution consists in a randomized search for the maximum likelihood estimate combined with the Metropolis MCMC algorithm in the Bayes framework.

[1]Department of Stochastic Informatics, ÚTIA AV ČR, Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic, volf@utia.cas.cz

2 Competing risks scheme

Let us recall briefly the competing risks scheme: There are K (possibly dependent) random variables, times $T_j, j = 1, ..., K$, running simultaneously. Observation is terminated at minimum of them. Sometimes there is another random or deterministic variable C of right censoring assumed to be independent of all T_j. Standardly, C is the time of observation termination without expected event occurrence; in that case $C < T_j$ for all j. We assume that all variables T_j are of continuous type. Let $\overline{F}_K(t_1, ..., t_K) = P(T_1 > t_1, ..., T_K > t_K)$ be the joint survival function of $\{T_j\}$. However, instead the 'net' survivals T_j we observe just 'crude' data (sometimes called also 'the identified minimum') $Z = \min(T_1, ..., T_K, C)$ and the indicator $\delta = j$ if $Z = T_j$, $\delta = 0$ if $Z = C$. Such data lead us to direct estimation of the distribution of $Z^* = \min(T_1, ..., T_K)$, for instance its survival function $S(t) = P(Z^* > t) = \overline{F}_K(t, ..., t)$. Further, we can estimate so called **incidence densities**

$$f_j^*(t) = dP(Z^* = t, \delta = j) = -\frac{\partial \overline{F}_K(t_1, ..., t_K)}{\partial t_j} | (t_1 = ... = t_K = t),$$

and also their integrals, cumulative incidence functions

$$F_j^*(t) = \int_0^t f_j^*(s)\, ds = P(Z^* \le t, \delta = j).$$

Notice that $\lim F_j^*(t) = P(\delta = j) < 1$ if $t \to \infty$, $S(t) = 1 - \sum_{j=1}^K F_j^*(t)$. Further, so called cause–specific hazard functions for events $j = 1, 2, \ldots, K$ are estimable by:

$$h_j^*(t) = \lim_{d \to 0} \frac{P(t \le Z^* < t + d,\, \delta = j \,|\, Z^* \ge t)}{d}.$$

Overall hazard rate for $Z^* = \min(T_1, ..., T_K)$ is then:

$$h^*(t) = \lim_{d \to 0} \frac{P(t \le Z^* < t + d \,|\, Z^* \ge t)}{d} = \sum_{j=1}^K h_j^*(t),$$

by integration the cumulated hazard rates $H_j^*(t)$, $H^*(t)$ are obtained. Consequently, $S(t) = P(Z^* > t) = \exp(-H^*(t))$. Then $f_j^*(t) = h_j^*(t) \cdot S(t)$ and the cumulative incidence functions can be also written as

$$F_j^*(t) = P(Z^* \le t, \delta = j) = \int_0^t S(s) \cdot h_j^*(s)\, \mathrm{d}s.$$

As both components, i.e. S and h_j^*, are estimable consistently by standard survival analysis methods, it follows that there also exist consistent estimates of F_j^*, see for instance Lin (1997).

2.1 Non-identifiability problem

As it has already been said, in general, from data (Z_i, δ_i), $i = 1, \ldots, N$ it is not possible to identify neither marginal nor joint distribution of $\{T_j\}$. A. Tsiatis (1975) has shown that for arbitrary joint model we can find a model with independent components having the same incidences, i.e. we cannot distinguish among the models. Namely, this 'independent' model is given by cause-specific hazard functions $h_j^*(t)$. It follows that it is necessary to make certain functional assumptions about the form of both marginal and joint distribution in order to identify them. Several such cases are specified for instance in Basu and Ghosh (1978). More recent results on identifiability can be found for example in Schwarz et al (2013) dealing with non-parametric setting, or in Escarela and Carriere (2003) considering Frank copula and parametric models.

Many authors have studied the role of additional information gathered from covariates in the framework of a regression model for examined random times. There are numerous results showing conditions for full identifiability of such a regression model, starting from already mentioned Heckman and Honoré (1989). Lee (2006) investigated more general transformation models of regression. Berg et al. (2007) have studied two competing transition rates from unemployment state. They have used a discrete-time multiplicative regression model with latent heterogeneities. A common identifiability assumptions consists in sufficiently rich structure of covariates. However, all these studies rely on the assumption that the dependence structure (in the next section given by a copula parameter) does not change with covariates. And this rater strong assumption is the target of our examination.

2.2 Copulas in models for competing risks

Let us reduce the model to just 2 competing events, random variables S, T (eventually with a censoring variable C). The data are then given as realizations of N i.i.d. random variables $Z_i = min(S_i, T_i, C_i), \delta_i = 1, 2, 0$, $i =$

$1, 2, ..., N$. The notion of copula offers a way how to model multivariate distributions, here the joint distribution function $F_2(s,t)$ of S, T:

$$F_2(s,t) = \boldsymbol{C}(F_S(s),\, F_T(t),\, \theta), \tag{1}$$

F_S, F_T are marginal distribution functions of S, T, $\boldsymbol{C}(u,v,\theta)$ is a copula, i.e. a two-dimensional distribution function on $[0,1]^2$, with uniformly on $[0,1]$ distributed marginals U, V, θ is a copula parameter. The parameter is connected uniquely with correlation of U, V, hence also with correlation of S, T. It is seen that the use of copula allows to model the dependence structure separately from the analysis of marginal distributions. Hence, the identifiability of the copula (and its parameter) and marginals can be considered as two separate steps.

Zheng and Klein (1995) have proved that when the copula is known, the marginal distributions are estimable consistently (and then the joint distribution, too, from (1)), even in non-parametric (so that quite general) setting. However, in general, also value of θ is needed, because (again due to Tsiatis, 1975) without fully determined copula we are not able to distinguish between the 'true' model and corresponding independent one. On the other hand, Zheng and Klein (1995) also argued that the selection of copula type is not crucial. The problem of proper copula choice is analyzed in a set of papers, let us mention here Kaishev et al (2007) comparing performance of several copula types. A common agreement is that the knowledge (or a good estimate) of parameter θ is much more important for correct model of joint distribution.

As a consequence, because the knowledge of copula type is still an unrealistic supposition, we can try to use certain sufficiently flexible class of copulas, as approximation, and concentrate to reliable estimation of its parameter. There exist a large number of different copula functions, among them for instance a set of Archimedean copulas. Let us concentrate here to one rather universal and flexible copula type, namely to Gauss copula.

3 Use of Gauss copula

Let X, Y be standard normal random variables $\sim N(0,1)$ tied with (Pearson) correlation $\rho = \rho(X,Y)$. Let us denote by $\varphi(x)$, $\phi(x)$ the univariate standard normal density and distribution function, further by $\phi_2(x,y)$ distribution function and by $\varphi_2(x,y)$ density function of two-dimensional Gauss distribution with both expectations equal to zero and covariance matrix $\Sigma = [1, \rho; \rho, 1]$. If we define $U = \phi(X)$, $V = \phi(Y)$, we obtain that the couple (U,V) has a 2-dimensional copula distribution on $(0,1)^2$ with distribution function

$$\boldsymbol{C}(u,v) = \phi_2\big(\phi^{-1}(u),\, \phi^{-1}(v)\big). \tag{2}$$

Naturally, $\rho(U,V) \neq \rho(X,Y)$ (though they are rather close, as a rule), while Spearman's correlations coincide, namely $\rho_{\mathrm{SP}}(X,Y) = \rho_{\mathrm{SP}}(U,V) = \rho(U,V)$. As our aim is to model the dependence of competing variables S, T, let us assume that their joint distribution function is given by Gauss copula (2),

$$F_2(s,t) = \phi_2\big(\phi^{-1}(F_S(s)),\, \phi^{-1}(F_T(t))\big), \tag{3}$$

and $S = F_S^{-1}(\phi(X))$, $T = F_T^{-1}(\phi(Y))$. Again $\rho_{\mathrm{SP}}(S,T) = \rho_{\mathrm{SP}}(U,V)$, and "initial" $\rho = \rho(X,Y)$ is the only parameter describing the dependence of S and T. It, naturally, differs from $\rho(S,T)$, however, all values $\rho(S,T)$ can be achieved by convenient choice of $\rho(X,Y)$. Let us remark here that the real dependence among S, T can be much more complicated, nevertheless the use of Gauss copula offers here certain rather simple and sufficiently flexible (as regards the correlation) set of distributions.

3.1 Estimation in Gauss copula model

When parameter ρ is known, copula (2) is fully defined and from Zheng, Klein (1995) it follows that the distribution of (S,T) can be estimated, even non-parametrically. On the other hand, without knowledge of ρ nonparametric model is not identifiable and in the parametric setting explicit proofs of identifiability are available for just certain types of marginal distributions specified for instance already in Basu and Ghosh (1978). We shall deal with a richer model including the covariates, and, similarly like Han and Hausman (1990), with the Cox's model of dependence on them. Let us first sketch the estimation procedure based on the maximum likelihood method. The data are $(Z_i, \delta_i),\ i = 1, \ldots, N$, the likelihood function then has the form

$$L = \prod_{i=1}^{N} \left\{ -\frac{\partial}{\partial s} \overline{F}_2(s,t) \right\}^{I[\delta_i=1]} \cdot \left\{ -\frac{\partial}{\partial t} \overline{F}_2(s,t) \right\}^{I[\delta_i=2]} \cdot \overline{F}_2(s,t)^{I[\delta_i=0]},$$

evaluated at $s = t = Z_i$, with $\overline{F}_2(s,t) = P(S > s,\, T > t) = 1 - F_S(s) - F_T(t) + F_2(s,t)$. From (3) it follows that $F_2(s,t) = \phi_2(x,y)$ with $x = \phi^{-1}(F_S(s))$, $y = \phi^{-1}(F_T(t))$. For the first term we obtain that

$$\frac{\partial}{\partial s} \overline{F}_2(s,t) = -f_S(s) + \frac{\partial}{\partial s} \phi_2(\phi^{-1}(F_S(s)), \phi^{-1}(F_T(t))) =$$

$$= -f_S(s) + \frac{\partial}{\partial x}\phi_2(x,y) \cdot \frac{d\phi^{-1}(F_S(s))}{ds} = -f_S(s) + \phi_1(y; \rho x, 1-\rho^2) \cdot \varphi(x) \cdot \frac{f_S(s)}{\varphi(x)},$$

where $\phi_1(y; \mu, \sigma^2)$ denotes the distribution function of normal distribution $N(\mu, \sigma^2)$, evaluated at y. The second term of the likelihood can be processed in the same way, hence the likelihood can be expressed in the following manner:

$$L = \prod_{i=1}^{N} \left\{ f_S(Z_i) \left[1 - \phi_1(Y_i; \rho X_i, 1-\rho^2)\right] \right\}^{I[\delta_i=1]} \cdot \left\{ f_T(Z_i) \left[1 - \phi_1(X_i; \rho Y_i, 1-\rho^2)\right] \right\}^{I[\delta_i=2]} \cdot$$

$$\cdot \left\{1 - F_S(Z_i) - F_T(Z_i) + \phi_2(X_i, Y_i)\right\}^{I[\delta_i=0]}, \tag{4}$$

again with $X_i = \phi^{-1}(F_S(Z_i))$, $Y_i = \phi^{-1}(F_T(Z_i))$. Parameter ρ is hidden in ϕ_1 and in ϕ_2. Distributions of S and T are present both explicitly and also implicitly, in transformed X_i, Y_i. It is seen that the problem of maximization may be a difficult optimization task and has to be solved by a convenient numerical procedure. In the following real data example we search for the MLE of parameters. As it can be rather computationally involving, the search is performed with the aid of the MCMC method, in the Bayes approach framework starting from conveniently chosen uniform priors (c.f. Gamerman, 1997). Such method then allows to obtain estimated Bayes credibility intervals for model parameters.

4 Real data example

The data are taken from the Statlib database: http://lib.stat.cmu.edu/datasets/caseK.txt, the "Case K" data, appearing also in Kadane and Woodworth (2004). They did not use the idea of competing risks, the aim of their study was to explore whether older employees were or not "discriminated" having higher rate of dismissal. In fact, the same question is, though just implicitly, behind our analysis, the positive dependence of both risks revealed in the present paper can be interpreted also in such a way that under the risk of dismissal some people prefer to leave the job voluntarily, the change it in time. And the aim is to show that younger employees have higher tendency to do it. This phenomenon should be taken into account when comparing younger and older persons.

The data contain the records on all persons employed by a firm during the period of observation, from 1.1.1900 to 31.1.1995, namely their dates of birth, dates when persons were hired by the company and when they have left it, either voluntarily or were forced to leave (dismissed). There were together 412 people, from them 96 were fired, 108 left voluntarily, the rest, 208 employees, were still with the company at the end of data collection period. Hence, we deal with two competing risks of the end of work, and we assume that the risks are dependent. The time considered is the calendar time, in days, from 1 to 1857, the end of study is also the fixed time of censoring, namely $C = 1857$ is the upper bound for each personal record (it is so called type I censoring by fixed value). It is expected that the development of the company can be the reason for changing rates of leaving it. There are also people joining the company during the followed period, thus changing the "risk set" of the study.

The age (in years) of employees at the moment of leave or censoring was taken as a covariate, because it was expected that the age can influence the decision and is changing the relation between the rate of compulsory and voluntary leaving. The age varied from 20 to 70 years, its median was 39. Figure 1 shows the times of leaving, distinguishing both ways. It is seen that the period of higher intensity of dismissals (possibly as a consequence of certain problems of the company) started at about day 800 and lasted almost another 800 days, and was more remarkable for the strata with higher age, in fact supporting the conjecture of Kadane and Woodworth. In order to prove the dependence of correlation of random times of both competing events on age, the sample was divided into two subsamples with age$<$40 and age$\geq$40 years, each subsample was analyzed separately. Number of employees in each group and observed incidence is given in Table 1 below.

4.1 Competing risks with Cox's regression model

The competing risks model together with the Gauss copula was sketched in preceding section. The influence of the covariate (age) was incorporated via the Cox's regression model. Hence, each of both random times was described by the hazard rate

$$h_j(t; x) = h_{0j}(t) \cdot \exp(\beta_j \cdot x),$$

where $j = 1, 2$ for related risks of dismissal and voluntary leave, respectively, h_{0j} are baseline hazard rates and β_j are regression parameters. Covariate x is the age of persons (in years) at moment of job termination or of censoring, t is the time of study in days. Further, to make computation easier, we assume that the baseline distributions correspond to Weibull ones. The baseline hazard rates and their cumulated (integrated) versions have then the form

$$h_{0j}(t) = \frac{b_j \cdot t^{b_j - 1}}{a_j^{b_j}}, \quad H_{0j}(t) = \left(\frac{t}{a_j}\right)^{b_j},$$

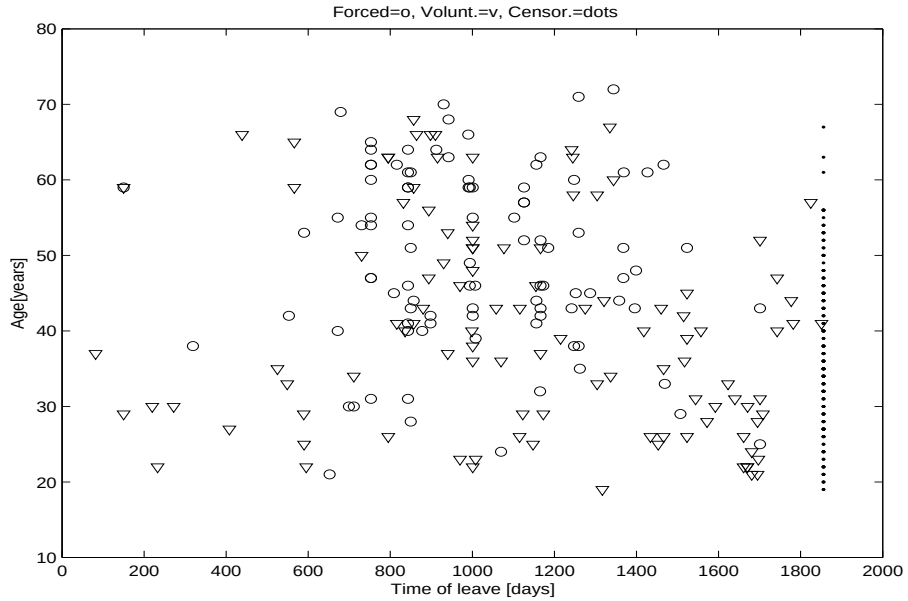


Figure 1 Graphical representation of data: circles=dismissals, triangles=voluntary leavings, dots=censored records.

where $a_j, b_j, j = 1, 2$ are Weibull scale and shape parameters. When the Cox's regression term is added, the distribution of both random times for given covariate value x is still Weibull, with the same shape parameters b_j and scale parameters depending on covariate, i.e.

$$a_j(x) = a_j \exp(-\frac{\beta_j}{b_j} x).$$

The complete model then contains 7 unknown parameters, a_j, b_j, β_j for $j = 1, 2$ and parameter ρ controlling the dependence of both competing risks via the Gauss copula.

4.2 Results

The model is fully parametrized, parameters were estimated by the MLE method. Already from (4) it is seen that the computation may be difficult, therefore the maximum of log-likelihood was found approximately with the aid of a random search. The results are collected in Table 1. In order to get also certain insight into the credibility of parameters values we have employed the Metropolis algorithm in the Bayes framework. Such a procedure yields the representation of the posterior distribution of parameters, hence also credibility intervals for them. Prior distributions of parameters were chosen uniform in reasonable intervals, the first rough estimate was obtained under the assumption of no dependence (i.e. $\rho = 0$). Then, naturally, the value of parameter ρ was alternated, too. Each computation used 10000 iterations of the algorithm, credibility intervals were obtained from the last half of them. It is seen that estimated parameters differ in both parts of data, i.e. they depend on the covariate - age of employees. It concerns also parameter ρ. The interpretation of higher positive correlation in the group of younger employees could be that they are more flexible and in the case of symptoms of approaching negative changes in the company they are more prone to search for a new employment. Further, it is seen that also Cox's model parameters β characterizing the dependence of hazard rates on age differ for both risks. For instance, from the first column of Table 1 we can deduce that the risk of dismissal increases significantly with age (parameter β_1). In fact, it could be understood as an indicator of existing discrimination of older employees in the sense of the study of Kadane and Woodworth (2004).

5 Conclusion

We have studied the problem of competing risks with regression, with the focus on assessing mutual dependence of competing random variables. The joint distribution was expressed with the aid of Gauss copula, while the Cox's regression model described the covariate influence. The model was utilized in an example with real unemployment data. Statistical analysis revealed positive correlation between times to both competing events, and also the dependence of correlation on the covariate. This was, in fact, the main purpose of the study. It has to be said that the Weibull model, used in order to simplify computations, is far from optimal. Further, the experience with present and similar computational procedures indicates that the log-likelihood function is, as a rule, rather flat,

	Data:	all data		with age< 40		with age≥ 40
Par.	$n_{1,2,0}$:	96, 108, 208		16, 53, 141		80, 55, 67
a_1	13723	(13150,15608)	4667	(3731,5920)	12165	(10418,14602)
b_1	1.975	(1.863,2.053)	1.444	(1.243,1.672)	2.070	(1.900,2.254)
a_2	3186	(3152,3420)	2817	(2582,3072)	8924	(6021,11395)
b_2	1.904	(1.776,1.910)	1.443	(1.300,1.617)	2.337	(2.117,2.597)
β_1	0.0713	(0.0652,0.0722)	0.0063	(0.0027,0.0120)	0.0717	(0.0616,0.0789)
β_2	0.0090	(0.0085,0.0100)	0.0002	(-0.0001,0.0003)	0.0600	(0.0507,0.0761)
ρ	0.221	(0.025,0.263)	0.778	(0.688,0.944)	0.216	(-0.073,0.580)

Table 1 Results of MCMC: Estimated parameters (modes of posterior distribution) and 90% credibility intervals. Scale parameters a_1, a_2 of Weibull baseline distributions are related to time in days; n_1, n_2, n_0 are numbers of people dismissed, leaving voluntarily, censored.

the convergence of computations to its maximum is slow, resulting confidence intervals are then quite wide. This phenomenon does not depend on the copula choice, it is a consequence of complicated model structure.

Acknowledgement: The research was supported by the grant No. 18-02739S of the Grant Agency of the Czech Republic.

References

[1] Basu, A.P. and Ghosh, J.K.: Identifiability of the Multinormal and Other Distributions under Competing Risks Model, *Journal of Multivariate Analysis* **8** (1978), 413–429.

[2] Van den Berg, G.J., van Lomwel, A.G.C., and van Ours, J.C.: Nonparametric estimation of a dependent competing risks model for unemployment durations, *Empirical Economics* **34** (2008), 477-491

[3] Escarela, G. and Carriere, J.F.: Fitting competing risks with an assumed copula, *Statistical Methods in Medical Research* **12** (2003), 333–349

[4] Gamerman, D.: *Markov Chain Monte Carlo*. Chapman and Hall, New York, 1997.

[5] Han, A. and Hausman J.A.: Flexible parametric estimation of duration and competing risk models, *J. of Applied Econometrics* **5** (1990), 1–28.

[6] Heckman, J.J. and Honoré, B.E.: The identifiability of the competing risks model, *Biometrika* **76** (1989), 325–330.

[7] Kaishev, V.K., Dimitrova, D.S., and Haberman, S.: Modelling the joint distribution of competing risks survival times using copula functions, *Insurance: Mathematics and Economics* **41**(2007), 339–361

[8] Kadane, J.B. and Woodworth, G.G.:Hierarchical Models for Employment Decisions, *Journal of Business and Economic Statistics* **22** (2004), 182–193

[9] Lee, S.: Identifcation of a competing risks model with unknown transformations of latent failure times, *Biometrika* **93** (2006), 996–1002.

[10] Lin, D.Y.: Non-parametric inference for cumulative incidence functions in competing risks studies, *Statistics in Medicine* **16** (1997), 901–910.

[11] Schwarz, M., Jongbloed, G., and Van Keilegom, I.: On the identifiability of copulas in bivariate competing risks models, *Canadian Journal of Statistics* **41** (2013), 291–303.

[12] Tsiatis, A.: A nonidentifiability aspects of the problem of competing risks, *Proc. Nat. Acad. Sci. USA* **72** (1975), 20–22.

[13] Zheng, M. and Klein, J.P.: Estimates of marginal survival for dependent competing risks based on an assumed copula, *Biometrika* **82** (1995), 127–138.

Measuring systemic risk in the European insurance sector using copula-DCC-GARCH model and selected clustering methods

Stanisław Wanat[1], Anna Denkowska[2]

Abstract. The subject of the present article is the study of correlations between large insurance companies and their contribution to systemic risk in the insurance sector. Our main goal is to analyse the conditional structure of the correlation on the European insurance market and to compare systemic risk in different regimes of this market. These regimes are identified by monitoring the weekly rates of returns of eight of the largest insurers (five from Europe and the biggest insurers from the USA, Canada and China) during the period February 2006 to March 2017. To this aim we use statistical clustering methods for time units (weeks) to which we assigned the conditional variances obtained from the estimated copula-DCC-GARCH model. In each of the identified market regimes we determined the commonly used today CoVaR systemic risk measure. From the analysis performed we conclude that all the considered insurance companies are positively correlated and this correlation is stronger in times of turbulences on global markets which shows an increased exposure of the European insurance sector to systemic risk during crisis. Moreover, in times of turbulences on global markets the value level of the CoVaR systemic risk index is much higher than in "normal conditions".

Keywords: systemic risk, insurance market, copula-DCC-GARCH.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Following the financial crisis in the years 2007-2009 and the European public debt crisis in the years 2010-2012, there has been significantly growing interest in systemic risk. This resulted in a prolific specialized literature proposing, among others, a wide range of new methods for the study of the influence of financial institutions on systemic risk. Moreover, both the academic community and the financial regulatory authorities paid more attention to the role played by non-bank financial institutions, in particular insurance companies, in creating systemic risk. Before the crisis it was generally accepted that the insurance market has a negligible impact on it. However, in the subsequent literature – although many a study still supported the latter point of view – there appeared several articles suggesting the possibility of the insurance market creating systemic risk. Let us quote here a few articles the authors of which claim that insurance companies:

- have become an unavoidable source of systemic risk (e.g. Billio et al. [6] , Weiß, Mühlnickel [21]);
- can be systemically important, but only due to their non-traditional (banking) activities (e.g. Baluch et al. [2], Cummins, Weiss [9]), while in general the systemic significance of the insurance sector as a whole is still subordinated to the banking sector (Chen et al.[8], Czerwińska [10]);
- are systemically unimportant due to the low level of interconnections and the lack of a strong dependance on external funding (e.g. Harrington [14], Bell, Keller [4], Geneva Association [13], Bednarczyk [3]).

On the other hand, in Bierth et al. [5] the authors, after having studied a very large sample of insurers in a long-term horizon, claim that the contribution of the insurance sector to systemic risk is relatively small, its peak having been reached during the financial crisis in the years 2007-2008. They also indicate the four L's: linkages, leverage, losses, liquidity, as the crucial factors influencing the exposure of insurers to systemic risk.

The present article belongs to the main stream of the studies concerning the linkages between large insurance companies and their contribution to systemic risk in the insurance sector. Our main aim is to check whether the strength of the existing connections between the eight largest insurers (five from Europe, one from the USA, Canada and China) together with their contribution to systemic risk in the European insurance sector depend on

[1] Cracow University of Economics, Department of Mathematics, 27 Rakowicka St., 31-510 Cracow, Poland e-mail: wanats@uek.krakow.pl.
[2] Cracow University of Economics, Department of Mathematics, 27 Rakowicka St., 31-510 Cracow, Poland e-mail: denkowsa@uek.krakow.pl..

the insurance market regime. The market regimes were identified by analysing weekly rates of return of the insurers in question during the period from February 2006 to March 2017. They were assessed using statistical clustering methods of time units (weeks) to which there had been assigned conditional variances obtained from the estimated copula-DCC-GARCH model. Indeed, we assumed that the change (increase) of the risk (variance) is a good (and, moreover, classical) index of the financial market tension. The advantage of such an approach is that there is no need to assume a priori a number of market regimes, because the latter is identified by the clustering quality assessment. Next, in each of the identified regimes we established the CoVaR systemic risk measure, commonly used nowadays (see e.g. Acharya et al. [1], Bierth et al. [5], Jobst [15]). We assumed that the European insurance market is represented by the weekly rates of return from the STOXX 600 Europe Insurance index. The CoVaR measure, indicating the contribution of each of the insurers to systemic risk, was assessed using the conditional distributions obtained from eight bivariate copula-DCC-GARCH models. In each of these models one boundary distribution represents the European insurance market (the logarithmic return from the STOXX 600 Europe Insurance index), while the other one represents the insurer (the appropriate logarithmic rate of return). To the best of our knowledge, such an approach has not been used in systemic risk analysis ever before.

The paper is organized as follows. The second chapter presents the methodology and the empirical strategy used in the paper, the third one presents the data and discusses the results obtained, while the fourth chapter contains the conclusions.

2 Methodology

The empirical strategy we use in this article in order to analyse the dependences and assess systemic risk on the European insurance market consists of two basic steps:

1. Market regime identification.
2. Analysis, in the identified market regimes, of:
   a) the dependencies between the studied insurance companies,
   b) the systemic risk.

In the first step it is assumed that market regimes are identified using statistical clustering methods of weekly periods *t* according to conditional variances of returns of all analysed instruments. The conditional variances that are essential in this approach are obtained through the multivariate copula-DDC-GARCH model. In this model the distribution of the rates of return vector $r_t = (r_{1,t},...,r_{k,t})'$, $t = 1,...,T$ conditional on the information set available at time $t-1$ (Ω_{t-1}) is modelled using conditional copulas introduced by Patton (2006). It takes the following form:

$$r_{1,t} \mid \Omega_{t-1} \sim F_{1,t}(\cdot \mid \Omega_{t-1}),...,r_{k,t} \mid \Omega_{t-1} \sim F_{k,t}(\cdot \mid \Omega_{t-1}) \tag{1}$$

$$r_t \mid \Omega_{t-1} \sim F_t(\cdot \mid \Omega_{t-1}) \tag{2}$$

$$F_t(r_t \mid \Omega_{t-1}) = C_t\left(F_{1,t}(r_{1,t} \mid \Omega_{t-1}),..., F_{k,t}(r_{k,t} \mid \Omega_{t-1})\right) \tag{3}$$

where C_t denotes the copula, whereas F_t and $F_{i,t}$ are the multivariate CDF and the CDFs of the marginal distributions at time *t*, respectively. The use of a copula allows the separate modelling of marginal distributions and the dependence structure of the vector r_t. In the empirical study elliptical copulas are used to describe the dynamics of the dependence structure, while conditional marginal distributions are modelled with the use of ARMA-GARCH models. Conditional correlation matrices R_t modelled with the use of DCC(1, 1) model (Engle [4]) are assumed to be the parameters of conditional copulas. The parameters are estimated using the inference function for margins - IFM approached. This method is presented in details e.g. in: Joe [16], p. 299–307, Doman [11], p. 35–37, Wanat [20], p. 98-99. The computations were done in the R environment using the ”rmgarch” package developed by Alexios Ghalanos.

To identify financial market regimes, statistical methods of unsupervised classification are used. The groups obtained are expected to be periods with a similar level of risk (i.e. similar conditional variance). Although the number of groups is not known a priori, it is assumed that it should be neither too small nor too large. In fact, clustering results are assessed taking into account both statistical criteria and economic interpretation of financial market regimes obtained. Clustering is conducted by means of hierarchical methods in which groups are created recursively by linking together the most similar objects (Ward's method is applied here). Other methods of division, i.e. the k-means method and the partitioning among medoids (PAM) method proposed by Kaufman and Rousseeuw [17] are also used. In both cases, after making the initial decision about the desired number of groups, objects are allocated in such a way that the relevant criterion is met. In order to evaluate the optimal number of

clusters in the data, we use internal validity indexes: Calinski-Harabasz pseudo F statistics (Calinski and Harabasz [7]), the average silhouette width (Kaufman and Rousseeuw [17]), the Dunn index (Dunn [12]), and Xie and Beni's [22] index. The final classification of objects is, therefore, the result of the comparison of the results of respective grouping algorithms.

In the second stage of analysis, in each of the identified market regimes we assessed the *CoVaR*. The systemic risk measure $CoVaR_{\beta,t}^{i|j}$ is defined to be the value at risk (*VaR*) of the institution (market index) *i* under the condition that another institution (market index) *j* is subject to distress i.e. its rate of return is smaller than its value at risk, meaning:

$$P\left(r_{i,t} \le CoVaR_{\beta,t}^{i|j} \mid r_{j,t} \le VaR_{\alpha,t}^{j}\right) = \beta\,, \tag{4}$$

where $VaR_{\alpha,t}^{j}$ is the value at risk of the institution (market index) *j* at time *t* and α, β denote the significance levels for VaR and CoVaR measures, respectively. Using the conditional probability formula we get:

$$\frac{P\left(r_{i,t} \le CoVaR_{\beta,t}^{i|j},\ r_{j,t} \le VaR_{\alpha,t}^{j}\right)}{P\left(r_{j,t} \le VaR_{\alpha,t}^{j}\right)} = \beta \tag{5}$$

The definition of the value at risk for the institution *j*, i.e. $VaR_{\alpha,t}^{j}$ yields $P\left(r_{j,t} \le VaR_{\alpha,t}^{j}\right) = \alpha$, that is:

$$P\left(r_{i,t} \le CoVaR_{\beta,t}^{i|j},\ r_{j,t} \le VaR_{\alpha,t}^{j}\right) = \alpha\beta \tag{6}$$

Therefore, the assessment of $CoVaR_{\beta,t}^{i|j}$ requires knowledge of the bivariate distribution F_t of the vector $\left(r_{i,t}, r_{j,t}\right)$. Due to the Sklar Theorem this distribution can be represented using the copula in the following way:

$$F_t\left(r_{i,t}, r_{j,t}\right) = C_t\left(F_i(r_{i,t}),\ F_j(r_{j,t})\right) \tag{7}$$

Invoking (4), $CoVaR_{\beta,t}^{i|j}$ can be determined by solving (numerically) the equation:

$$C_t\left(F_i(CoVaR_{\beta,t}^{i|j}),\ \alpha\right) = \alpha\beta \tag{7}$$

In the empirical analysis we studied the influence on the European insurance market's systemic risk of the five largest insurance companies from Europe and the biggest insurers from the USA, Canada and China. In accordance, it was assumed that $r_{i,t}$ represents the European insurance market (we made use of the weekly rates of return from STOXX 600 Europe Insurance), while $r_{j,t}$ describes the insurers (we made use of the weekly logarithmic returns on shares). For each of the eight pairs:

(*logarithmic return of the* STOXX 600 *index* $r_{i,t}$, *logarithmic return of the insurer* $r_{j,t}$)

we assessed the parameters of the bivariate copula-DCC-GARCH model. Then, using these parameters together with the conditional correlations obtained by these models, we determined the copulas C_t and the distributions F_t. The values $CoVaR_{\beta,t}^{i|j}$ for the analysed period were obtained by solving numerically the equation (7).

3 Data and results

The data used in this study consist of the weekly log-returns of the five largest insurers from Europe and the biggest insurers from the USA, Canada and China (Tab.1) as well as the STOXX 600 Europe Insurance index representing the European insurance market from February 2016 to March 2017.

In the first stage of study we identified the regimes of the insurance market on the basis of the conditional variances of the rates of return of the insurance companies in question. These were assessed using the 8-variate copula-DCC-GARCH model. During the analysis we considered various ARMA-GARCH specifications of univariate models. Finally, on the grounds of information criteria and model appropriateness tests (result available upon request to the author) we opted, for all the instruments, for the ARMA(1,1)-eGARCH(2,2) model with the skew Student distribution. The eGARCH means exponential GARCH model put forward by Nelson. During the

analysis of the dynamics of the dependences between the rates of return we considered the Gauss and Student copulas together with various specifications of the DCC model. As earlier, in the basis of information criteria we chose the Student copula with conditional correlation coefficients obtained from the DCC(1, 1) model and a constant shape parameter η.

Insurer	Acronym	Country	Total assets (in mld USD)
AXA	AXA	France	944.145
Allianz	Allianz	Germany	934.654
Prudential plc	Prud	Great Britain	578.149
Legal & General	Legal	Great Britain	574.901
Aviva	Aviva	Great Britain	541.188
Metlife	Metlife	USA	898.764
Manulife Financial	Manu	Canada	534.705
Ping An Insurance	Ping	China	802.975

Table 1 Insurance companies considered in the study with their acronyms used in the presentation of the results

Market regimes were identified by means of clustering weekly periods with respect to the conditional variances of the rates of return of the studied insurance companies. In this crucial – from the point of view of the whole study – step we considered various combinations of distance measures, clustering methods and number of classes. We determined from 2 to 6 groups using each of the following methods: Warda, k-means and PAM. Eventually, led by criteria of clustering quality (results available upon request to the author) we chose a division into two classes obtained using the method of k-means with the Euclidean distance (cf. Fig. 1). In this case Rousseeuw's Silhouette internal cluster quality index is equal to 0.855, Calinski-Harabasz index: 1153.277, Dunn index: 0.073 and Xie-Beni index: 1.289. We assumed thereafter that to different classes there correspond different market regimes.

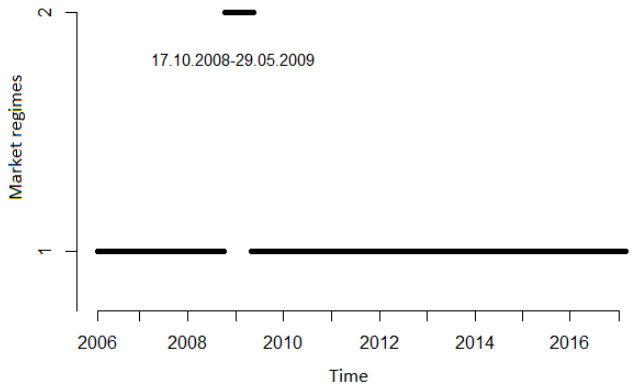


Figure 1 Identified market regimes

In the second step of our studies the analysis of the dependencies between the studied insurance companies was done based on the conditional correlations from the previously assessed 8-variate copula-DCC-GARCH model. Their distribution for the respective pairs in the identified market regimes is shown in Figure 2. The systemic risk assessment in the identified market regimes was performed using the CoVaR measure determined by the method described in the previous section. The CoVaR value distribution illustrating the influence of a given insurer on the European insurance market is shown in Figure 3.

4 Conclusions

In this work we used the copula-DCC-GARCH model to analyse the dependences in the group formed by the largest five insurance companies from Europe and the biggest insurers from the USA, Canada and China. Then, availing ourselves of the CoVaR measure we studied the influence of each insurer on the European insurance market systemic risk. The European market was represented by the STOXX 600 Europe Insurance index, while for the insurers we considered their quotations on domestic markets. The study was performed in two steps. The first one consisted in identifying the regimes of the European insurance market, while in the second one, we analysed – for the identified regimes – the following items: the correlations (using conditional correlations) between

the considered insurance companies, the dependences between a given insurer and the European insurance market as well as the influence of the studied insurance companies on the European insurance market systemic risk. The market regimes were identified by monitoring the insurers' logarithmic returns on shares. To this aim we applied statistical clustering methods for weekly periods to which we were assigning the conditional variances obtained from the estimated 8-variate copula-DCC-GARCH model. Both, the clustering quality measures and the possibility of a reasonable economical interpretation exposed two different market regimes in the considered period of time: a regime of low volatility (1st regime – "normal") and a regime of unstable quotations (2nd regime – "risk") that appeared during the time of the biggest turbulences experienced by the global markets.

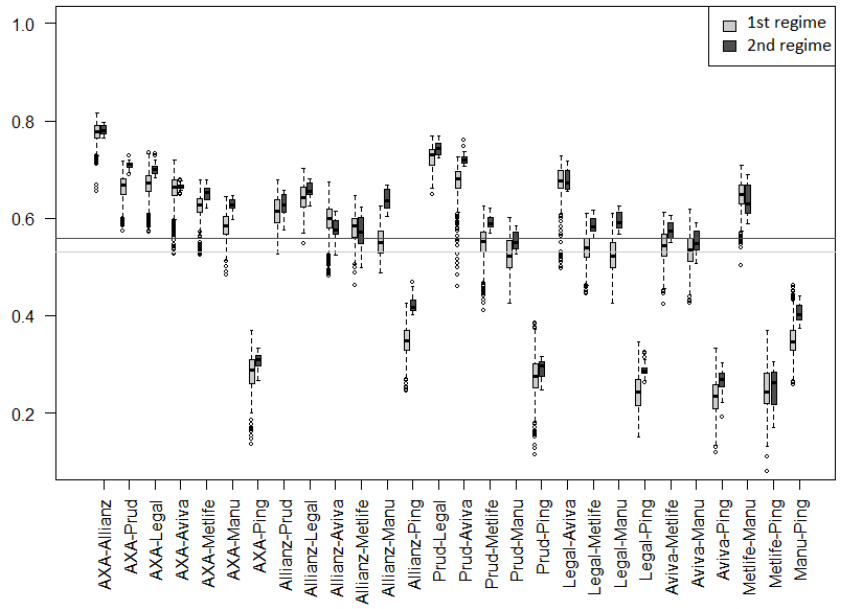


Figure 2 Distribution of the conditional correlations between analysed markets in the identified regimes

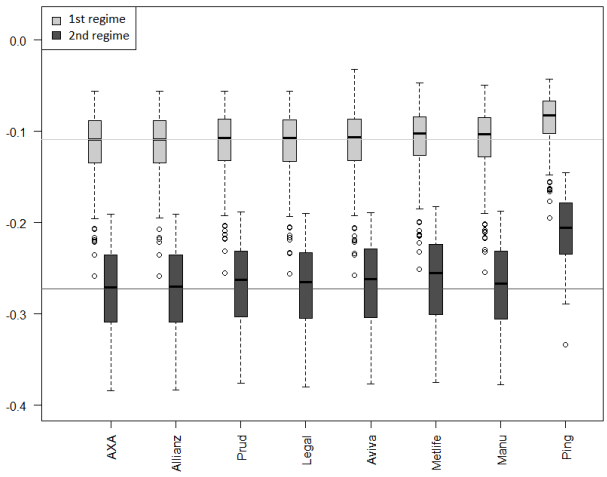


Figure 3 Distribution of the *CoVaR* measure in the identified regimes

We can draw the following conclusions from our study:

- The insurance companies from the investigated group are positively correlated. The strongest dependence is to be seen among the insurers from Europe (Axa and Allianz are the pair with the strongest tie), a somewhat weaker dependence exists between the insurers from Europe and those from North America, while the weakest is between the insurer from China and the remaining ones. These correlations are clearly stronger in the second identified regime, i.e. during the turbulences on global markets period (cf. Fig. 2). On that basis we can state that during a global crisis the exposure to systemic risk on the European insurance market increases.

- There is an important difference between the CoVaR measures for the first and second regimes of the European insurance market in the case of all the insurers from the studied group. The influence of insurance companies on systemic risk is much stronger during the turbulences period (cf. Fig. 3). It is also apparent that in a fixed regime this influence is more or less at the same level, which in the case of the insurer from China is somewhat lower than average.
- The influence of insurance companies from North America on the European insurance market systemic risk is at a comparative level with the influence of companies from Europe, both in the first and second identified market regimes.

Acknowledgements

The study is supported with subsidies for maintaining research capacity granted to the Faculty of Finance and Law at the Cracow University of Economics by the Polish Ministry of Science and Higher Education.

References

[1] Acharya, V. V., Pedersen, L. H., Philippon, T., Richardson, M. P. (2010), *Measuring Systemic Risk*, SSRN Electronic Journal, doi:10.2139/ssrn.1573171
[2] Baluch, F., Mutenga, S., Parsons, C. (2011), *Insurance, Systemic Risk and the Financial Crisis*, The Geneva Papers on Risk and Insurance - Issues and Practice,36 (1), 126-163. doi:10.1057/gpp.2010.40
[3] Bednarczyk,.T. (2013), *Czy sektor ubezpieczeniowy kreuje ryzyko systemowe?*, Studia Oeconomica Posnaniensia, 11 (260), 7-17
[4] Bell, M., Keller, B., (2009), *Insurance and Stability: The Reform of Insurance Regulation*, Zürich Financial Services Group Working Paper.
[5] Bierth, C., Irresberger, F., Weiß, G. N. (2015). *Systemic risk of insurers around the globe*. Journal of Banking & Finance, 55, 232-245. doi:10.1016/j.jbankfin.2015.02.014
[6] Billio, M., Getmansky, M., Lo, A. W., Pelizzon, L. (2012), *Econometric measures of connectedness and systemic risk in the finance and insurance sectors*, Journal of Financial Economics, 104 (3), 535-559, doi:10.1016/j.jfineco.2011.12.010
[7] Calinski, T., Harabasz, J., (1974), *A dendrite method for cluster analysis*, Communications in Statistics, 3 (1), 1–27
[8] Chen, H., Cummins, J. D., Viswanathan, K. S., Weiss, M. A. (2013), *Systemic Risk and the Interconnectedness Between Banks and Insurers: An Econometric Analysis*, Journal of Risk and Insurance, 81 (3), 623-652. doi:10.1111/j.1539-6975.2012.01503.x
[9] Cummins, J. D., Weiss, M. A. (2014), *Systemic Risk and The U.S. Insurance Sector*, Journal of Risk and Insurance, 81 (3), 489-528. doi:10.1111/jori.12039
[10] Czerwińska, T. (2014), *Systemic risk in the insurance sector,* Problemy Zarządzania, 12 (48), 41-63. doi:10.7172/1644-9584.48.3
[11] Doman, R., (2011), *Zastosowanie kopuli w modelowaniu dynamiki zależności na rynkach finansowych*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań
[12] Dunn, C. J., (1974), *Well-separated clusters and optimal fuzzy partitions*, Journal of Cybernetics 4, 95–104
[13] Geneva Association, (2010), *Systemic risk in insurance: an analysis of insurance and financial stability*, Technical report, Special Report of The Geneva Association Systemic Risk Working Group, Switzerland
[14] Harrington, S. E. (2009), *The Financial Crisis, Systemic Risk, and the Future of Insurance Regulation*, Journal of Risk and Insurance, 76 (4), 785-819. doi:10.1111/j.1539-6975.2009.01330.x
[15] Jobst, A. A. (2014), *Systemic Risk in the Insurance Sector: A Review of Current Assessment Approaches*, The Geneva Papers on Risk and Insurance - Issues and Practice, 39 (3), 440-470. doi:10.1057/gpp.2013.7
[16] Joe, H., (1997), Multivariate models and dependence concepts, Chapman-Hall, London
[17] Kaufman, L., Rousseeuw, P. J., (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: Wiley &Sons
[18] Nelson, D. B., (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, Econometrica, 59, 347–370
[19] Patton, A. J., (2006), Modelling asymmetric exchange rate, International Economic Review, 47, 527–556
[20] Wanat, S., (2012), *Modele zależności w agregacji ryzyka ubezpieczyciela*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków
[21] Weiß, G. N., Mühlnickel J., (2014), *Why do some insurers become systemically relevant?*, Journal of Financial Stability, 13, 95-117
[22] Xie, X., Beni, G., (1991), *Validity measure for fuzzy clustering*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13, 841–847

The use of tests LRuc, LRind, LRcc on example of estimation of the Value-at-Risk for WIG, DAX and DJIA indices

Marek Zinecker[1], Michał Bernard Pietrzak[2], Marcin Fałdziński[3], Adam P. Balcerzak[4], Tomáš Meluzín[5]

Abstract. The subject matter discussed in the article concerns the application of the Value at Risk metric in risk measurement on capital markets. On the one hand, this risk is the result of continuous and dynamic growth of the network of interdependencies between financial markets, and, on the other hand, is the result of the occurrence of shock situations that may turn into a permanent crisis situation. An unexpected and significant increase in risk may, in turn, translate into significant losses of financial institutions, and in some extreme cases lead even to their collapse. Therefore, measuring market risk in a skilful manner provides the opportunity to protect against significant financial losses. Quantitative determination of market risk is also important due to the possibility of spreading financial markets shocks to real economies. The main research objective of this article is to assess the quality of the Value at Risk calculations performed for the capital markets of the United States, Germany, and Poland. The research was conducted in the time period 2000-2012, where the parameters of the DCC-GARCH model were estimated for the purpose of determining VaR. The assessment of quality was made based on the backtesting performed, where binominal tests were used and they took the form of the LRuc test, LRind test, and LRcc test.

Keywords: value-at-risk, backtesting, DCC-GARCH model, binominal test.

JEL classification: G15, C58
AMS classification: 91B30, 91B84

1 Introduction

Analysis of financial markets performed at the international level leads to the conclusion of continuous and dynamic growth of the network of interdependencies between these markets. A continuous increase in the value of financial markets is also observed, with a significant part of the share resulting from the increase in the capitalization of capital markets [see 25; 26; 11]. The strongest interdependencies are observed for capital markets, where the impact of information can be seen on subsequent markets almost in real time. Abundant research on the functioning of capital markets has indicated the existence of interdependencies between markets, where their strength changes over time. The problem of changing the strength of interdependence is interesting, since under a crisis situation the strength of interdependence tend to increase significantly, while periods of tranquillity on markets are characterized by a substantially lower level of mutual dependence [16; 17; 30; 31; 37; 39; 8]. Therefore, the problem of measuring market risk is gaining particular importance, since under a crisis situation it may be significantly increased [2; 27; 34; 21]. An unexpected and significant increase in risk may, in turn, translate into significant losses made by financial institutions, and in extreme cases it may lead even to their collapse. Quantitative determination of market risk is also important due to the possibility of spreading financial markets shocks to real economies. Additionally, changes in the real sphere concerning macro and micro economic development [28; 29; 12; 33], including the level of innovation and the technological potential [4; 40] and the level of investment, which affects economy international competitiveness [19] can be influenced by the situation on the financial markets, their growth dynamics and finally their effectiveness [5; 22; 23; 24; 20; 32].

Value at Risk that is recommended by financial supervision institutions is a market risk measurement technique commonly used by financial institutions [6; 3; 7; 15; 22]. The issues undertaken in this article focus on

[1] doc. ing. Marek Zinecker, PhD, Brno University of Technology, *Kolejní 2906/4, 61200 Brno, Czech Republic* e-mail: zinecker@fbm.vutbr.cz.
[2] Michał Bernard Pietrzak PhD, Nicolaus Copernicus University, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: michal.pietrzak@umk.pl
[3] Marcin Fałdziński PhD, Nicolaus Copernicus University, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: marf@umk.pl.
[4] Adam P. Balcerzak, PhD, Nicolaus Copernicus University, ul. Gagarina 13a, 87-100 Toruń, Poland, e-mail: adam.balcerzak@umk.pl.
[5] doc. ing. Tomáš Meluzín, PHD, Brno University of Technology, *Kolejní 2906/4, 61200 Brno, Czech Republic*, e-mail: meluzint@fbm.vutbr.cz.

the application of VaR in measuring risk on the capital markets of the United States, Germany, and Poland. The main research objective of the article is to assess the quality of the VaR calculations made for the major market indexes: DJIA, DAX and WIG. The assessment of quality was made based on the backtesting performed, where binominal tests were used in the form of the LRuc test, LRind test, and LRcc test. The research was conducted in the time period 2000-2012, where the parameters of the DCC-GARCH model were estimated for the purpose of determining VaR.

2 The DCC-GARCH model and Value at Risk

The Value at Risk (VaR) metric is commonly applied to measure the market risk of an individual financial asset or investment portfolio. By means of this metric, we can determine the probability of incurring a financial loss that would be considered as significant from the institution's point of view. To determine VaR for an asset or an investment portfolio, it is necessary to determine the distribution of their returns. In this case, the occurrence of returns variance that is volatile over time becomes problematic. The use of the multi-dimensional DCC-GARCH models allows for the modelling of the properties of returns in the form of the concentration of periods of significantly increased volatility and periods of reduced volatility in capital markets [see 13; 14; 35; 38; 18]. An additional advantage of the multidimensional DCC-GARCH models is their capacity to examine the mutual dependence between markets by estimating the variable conditional correlation. The DCC-GARCH model is determined as

$$\mathbf{Y_t} = \boldsymbol{\mu_t} + \boldsymbol{\eta_t},\ \boldsymbol{\eta_t}|F_{t-1}\sim t(0,\mathbf{H_t}),\ \mathbf{H_t} = \mathbf{D_t}\mathbf{R_t}\mathbf{D_t},\ \mathbf{D_t^2} = diag\{\mathbf{H_t}\}, \tag{1}$$

$$H_{i,t} = \omega_i + \alpha_i\eta_{i,t-1}^2 + \beta_i H_{i,t-1}, \boldsymbol{\varepsilon_t} = \mathbf{D_t^{-1}}\boldsymbol{\eta_t}, \tag{2}$$

$$\mathbf{R_t} = diag\{\mathbf{Q_t}\}^{-1/2}\mathbf{Q_t}diag\{\mathbf{Q_t}\}^{-1/2}, \tag{3}$$

$$\mathbf{Q_t} = \boldsymbol{\Omega} + a\boldsymbol{\varepsilon_{t-1}}\boldsymbol{\varepsilon'_{t-1}} + b\mathbf{Q_{t-1}},\ \boldsymbol{\Omega} = \bar{\mathbf{R}}(1-a-b), \tag{4}$$

where $\mathbf{Y}_t$ - multivariate process of returns, t - conditional t-distribution with $\nu > 2$ degrees of freedom, $\boldsymbol{\mu}_t$ - the vector of conditional means of returns, $\mathbf{H}_t$ - the conditional covariance matrix, $\mu_{i,t}$ - *i*-th equation of the conditional mean, $h_{i,t}$ - the conditional variance equation for *i*-th returns, where $i = 1,\ldots,N$, $\omega_i, \alpha_i, \beta_i$ - the parameters of the conditional variance equation, $\mathbf{R}_t$ - conditional correlation matrix, $\mathbf{Q}_t$ - quasi correlation matrix, $\overline{\mathbf{R}}$ - the unconditional covariance matrix of the standardize errors which can be estimated or set as $\bar{\mathbf{R}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon'_t}$, a, b the parameters of the conditional correlation equation.

Aielli [1] shows that the estimation of $\bar{\mathbf{R}}$ as the the empirical covariance matrix is inconsistent. So, we decided to estimate the unconditional covariance matrix.

Value at Risk allows us to determine the potential loss for the current market value of the possessed asset or portfolio. Together with the loss value, the probability of occurrence or exceeding of the determined loss at a given tolerance level is also determined. VaR determination consists in adopting a tolerance level. The VaR measure can be obtained using the following formula (3) [see 10]

$$\boldsymbol{-VaR} = \mathbf{C_\alpha} - \mathbf{C_0} \tag{4}$$

where $\mathbf{C_0}$ is the initial value of the asset (portfolio), $\mathbf{C_\alpha}$ is the quantile of the distribution of the asset (portfolio) value at the given tolerance level $\boldsymbol{\alpha}$ representing the final value of the asset (portfolio).

In practice, estimation of VaR consists in using the distribution of returns of the analysed asset (portfolio). The VaR estimate is then determined based on formula 5.

$$VaR_q^t = -\mu_{t+1} + \sqrt{h_{t+1}}z_q \tag{5}$$

where μ_{t+1} is a one period ahead forecast of conditional mean, h_{t+1} is a one period ahead forecast of conditional volatility and z_q is the *q*-quantile of conditional distribution. The DCC-GARCH model estimation is performed to determine the parameters of the distribution of returns.

An important problem is the assessment of quality of the VaR estimates. The VaR measure is tested for two properties: unconditional coverage and independence property. The first one assumes that the expected number of hits (cases when returns are greater than estimated VaR) given the assumed coverage level. The second one states that the hit process is independent. The most common approach to assessing the quality of VaR estimates is called backtesting, which is based on the examination of hits resulting from the estimated values of the

measure. In order to evaluate the quality of the VaR, the backtesting procedure is applied. In the article, for the purposes of the verification of the quality of VaR, binominal tests were used and they took the form of the unconditional coverage test LRuc, the test of independent LRind, and the conditional coverage test LRcc [see 9].

3 Empirical estimation of Value at Risk and the use of binominal tests

The study used the time series of three stock indices: DAX, DJIA and WIG, which allowed the analysis of the capital markets of Germany, the United States, and Poland to be carried out. Logarithmic daily returns were used for calculations. Data covering the time period 2000-2012 were obtained from the www.finance.yahoo.com website. In the first step, GARCH model estimation for individual indices was made, which allowed us to determine the best model specification in the form of GARCH (1,1).

Next, the parameters of the DCC-GARCH (1,1) model were estimated using the maximum likelihood method with the conditional *t*-distribution. The results of the estimation are presented in Table 1. All parameters, derived from both the conditional variance equations and conditional correlation equations, proved to be statistically significant at the 5% significance level, beside the unconditional correlation parameter $\bar{R}_{23}$. The occurrence of thick tails in the distribution of returns was confirmed by the estimates of *v* parameter of the *t*-distribution.

The conditional variance equations			
Parameter	**Estimate**	**Std. error**	**p-value**
ω_1 (WIG)	0.0197	0.0048	0.0001
α_1 (WIG)	0.0704	0.0077	0.0000
β_1 (WIG)	0.9208	0.0082	0.0000
ω_2 (DAX)	0.0123	0.0029	0.0000
α_2 (DAX)	0.0708	0.0081	0.0000
β_2 (DAX)	0.9188	0.0087	0.0000
ω_3 (DJIA)	0.0176	0.0048	0.0003
α_3 (DJIA)	0.0508	0.0068	0.0000
β_3 (DJIA)	0.9388	0.0080	0.0000
v	8.9814	0.7292	0.0000
The conditional correlation equation			
Parameter	**Estimate**	**Std. error**	**p-value**
a	0.0080	0.0020	0.0001
b	0.9919	0.0021	0.0000
$\bar{R}_{12}$	0.2739	0.1208	0.0234
$\bar{R}_{13}$	0.2749	0.1101	0.0126
$\bar{R}_{23}$	-0.0286	0.1241	0.8177

Table 1 The results of the estimation of the DCC-GARCH

The correct statistical properties of the estimated DCC-GARCH(1,1) model indicate the possibility of using this model to estimate Value at Risk pertaining to with the formula (4). The VaR estimation was started by estimating the DCC-GARCH model based on the first 1000 observations obtained from the set of returns for the time period 2000-2012 (3000 observations). The parameter estimates allowed us to compute a one-day ahead VaR forecast. Then, the procedure was repeated 1999 times, shifting each 1000 observations window by one day (the so called rolling window approach). The performed procedure led to 2000 VaR estimates for each market index separately.

The next step was to assess the quality of Value at Risk estimations for selected indices. For this purpose, the backtesting procedure was carried out, where statistical tests the LRuc test, LRind test, and LRcc test were applied. The results obtained are contained in Table 2. Based on the received test results, it can be concluded that for DJIA and DAX, VaR properties can be deemed to be correct. For the WIG index, we observe rejection of the null hypothesis for LRuc, LRind and LRcc tests. This means that VaR estimates are not in line with assumed properties.

Binominal tests		**LRuc**			**LRind**			**LRcc**	
coverage level	99%	95%	90%	99%	95%	90%	99%	95%	90%
				DJIA					
statistics	3.5917	1.2317	0.0498	0.8538	0.2615	0.4255	4.4455	1.4932	0.4753
p-value	0.0581	0.2671	0.8234	0.3555	0.6091	0.5142	0.1083	0.4740	0.7885
				DAX					
statistics	1.6611	5.2050	1.9517	0.6853	0.2920	2.4130	2.3464	5.4971	4.3646
p-value	0.1974	0.0225	0.1624	0.4078	0.5889	0.1203	0.3094	0.0640	0.1128
				WIG					
statistics	12.9450	0.0418	0.0502	1.6077	7.4608	6.2967	14.5527	7.5027	6.3469
p-value	**0.0003**	0.8379	0.8227	0.2048	**0.0063**	**0.0121**	**0.0007**	**0.0235**	**0.0419**

Note: Bold values means rejection of the null based on the 5% significance level. Source: own computations.

Table 2 Backtesting results

4 Conclusions

The subject matter discussed in the article concerned the application of the VaR metric to measure risk on the capital markets of Poland, Germany and the United States. This risk results from a dynamic increase in interdependencies between financial markets. Therefore, shock events occurring on the markets can contribute to a permanent crisis situation and generate significant financial losses. Therefore, quantifying market risk using Value at Risk is nowadays a significant research problem.

Following the research objective undertaken in this article, we estimated Value at Risk for the selected market indices, and then we evaluated their quality using binominal tests. The evaluation of quality was made using the backtesting procedure, where the LRuc test, LRind test, and LRcc test were used. The results obtained allowed us to determine the correct properties for VaR estimates for all of the analysed indices.

References

[1] Aielli, G. (2009). *Dynamic conditional correlations: on properties and estimation*. Department of Statistics, University of Florence, Mimeo.

[2] Akbalik, M. & Tunay, K. (2016). An analysis of Ramadan effect by GJR-GARCH model: case of Borsa Istanbul. *Oeconomia Copernicana*, 7(4), 593-612.

[3] Angelidis, T. & Degiannakis, S. (2006). Backtesting VaR Models: An Expected Shortfall Approach. *Working Papers, University of Crete, Athens University of Economics and Business.*

[4] Balcerzak, A. P. & Pietrzak, M. B. (2017). Digital Economy in Visegrad Coutnries. Multiple-criteria decision analysis at Regional Level in the Years 2012 and 2015. *Journal of Competitiveness*, 9(2), 5-18.

[5] Balcerzak, A. P., Kliestik, T., Streimikiene, D. & Smrčka L. (2017). Non-parametric approach to measuring the efficiency of banking sectors in European Union Countries. *Acta Polytechnica Hungarica*, 14(7), 51-70.

[6] Blanco, C. & Ihle, G. (1998). How Good Is Your VaR? Using Backtesting to Assess System Performance. *Financial Engineering News*, August, 1-2.

[7] Candelon, B., Colletaz, G., Hurlin, C. & Tokpavi, S. (2011). Backtesting Value-at-Risk: A GMM Duration-based Test. *Journal of Financial Econometrics*, 9(2), 314-343.

[8] Cevik, E.I., Korkmaz, T. & Cevik, E. (2017). Testing causal relation among Central and Eastern European equity markets: evidence from asymmetric causality test. *Economic Research-Ekonomska Istraživanja*, 30(1), 381-393.

[9] Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4), 841-862.

[10] Dowd, K. (2005). *Measuring Market Risk*. John Wiley & Sons Ltd.

[11] Dusatkova, M.S., Zinecker, M. & Meluzin, T. (2017). Institutional Determinants of Private Equity Market in Czech Republic. *Economics & Sociology*, 10(4), 83-98.

[12] Dvorsky, J., Sopkova, G. and Janoskova, M. (2017). Evaluation of Social Environment and Access to Financial Resources for Business: Case Study of the Czech and Slovak Republic. *Economicko-manazerskie spektrum/Economic and Managerial Spectrum*, 11(1), 62-73.

[13] Engle, R.F. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics*, 20, 339-350.

[14] Engle, R.F. (2009). *Anticipating Correlations A New Paradigm for Risk Management*. Princeton and Oxford: Princeton University Press.

[15] Fałdziński M. (2017). Backtesting value-at-risk for multiple risk levels: a lexicographic ordering approach. In: P. Pražák (ed.). *Mathematical methods in economics 2017: proceedings of the 35th International Conference*, 13th-15th September 2017, Hradec Králové: University of Hradec Králové, 155-160.

[16] Fałdziński, M., Balcerzak, A.P., Meluzín, T., Pietrzak, M. B. & Zinecker, M. (2016). Cointegration of Interdependencies Among Capital Markets of Chosen Visegrad Countries and Germany In A. Kocourek and M. Vavrousek (Eds.). *34th International Conference Mathematical Methods in Economics MME 2016 Conference Proceedings*. Liberec: Technical University of Liberec, 189-194.

[17] Heryán, T. & Ziegelbauer, J. (2016). 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), 61-74.

[18] Khan, F., Rehman, S.-U., Khan, H. & Xu, T. (2016). Pricing of risk and volatility dynamics on an emerging stock market: evidence from both aggregate and disaggregate data. *Economic Research-Ekonomska Istraživanja*, 29(1), 799-815.

[19] Kruk, H. & Waśniewska, A. (2017). Application of the Perkal method for assessing competitiveness of the countries of Central and Eastern Europe. *Oeconomia Copernicana*, 8(3), 337–352.

[20] Kubiszewska, K. (2017). Banking concentration in the Baltic and Western Balkan states — selected issues. *Oeconomia Copernicana*, 8(1), 65-82.

[21] Maknickiene, N., Lapinskaite, I. & Maknickas, A. (2018). Application of ensemble of recurrent neural networks for forecasting of stock market sentiments. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 13(1), 7-27.

[22] Meluzín, T., Balcerzak, A.P., Pietrzak, M. B., Zinecker, M., & Doubravský, K. (2018). The impact of rumours related to political and macroeconomic uncertainty on IPO success: evidence from a qualitative model. *Transformations in Business & Economics*, 17, 2(44), 148-169.

[23] Meluzín, T., Pietrzak, M. B., Balcerzak, A. P., Zinecker, M., Doubravský, K., & Dohnal, M. (2017). Rumours Related to Political Instability and their Impact on IPOs: The Use of Qualitative Modeling with Incomplete Knowledge. *Polish Journal of Management Studies*, 16(2), 171-187.

[24] Meluzín, T., Zinecker, M., Balcerzak, A.P., Doubravský, K., Pietrzak, M. B., & Dohnal, M. (2018). The timing of initial public offerings – non-numerical model based on qualitative trends. *Journal of Business Economics and Management*, 19(1), 63-79.

[25] Meluzin, T. & Zinecker, M. (2014). Research into Determinants Influencing 100 Decisions on the Polish Capital Market. *Betriebswirtschaftliche Forschung und Praxis*, 66(6), 652-671.

[26] Meluzin, T. & Zinecker, M. (2016). Trends in IPOS: the evidence from CEE capital markets. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 11(2), 327-341.

[27] Navid, B.J. & Shabantaheri, P. (2017). The development of capital market and bank risk: The Case of Iran, *International Review*, 1-2, 99-107.

[28] Pietrucha, J., Żelazny, R., Kozłowska, M. & Sojka, O. (2018). Import and FDI as channels of international TFP spillovers. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *13*(1), 55-72.

[29] Pietrzak, M. B., Balcerzak, A. P., Gajdos, A. & Arendt, Ł (2017). Entrepreneurial environment at regional level: the case of Polish path towards sustainable socio-economic development. *Entrepreneurship and Sustainability Issues*, *5*(2), 190-203.

[30] Pietrzak, M. B., Fałdziński, M., Balcerzak, A. P., Meluzín, T. & Zinecker, M. (2017). Short-term Shocks and Long-term Relationships of Interdependencies Among Central European Capital Markets. *Economics & Sociology*, 10(1), 61-77.

[31] Pietrzak, M. B., Łaszkiewicz, E., Balcerzak, A. P., Meluzín, T. & Zincker, M. (2017). Identification of the Direction of Changes in the Structure of Interdependence among the US Capital Market and the Leading European Markets In P. Pražák (Ed.). *35th International Conference Mathematical Methods in Economics MME 2017 Conference Proceedings.* Hradec Králové: University of Hradec Králové, 572-577.

[32] Skvarciany, V., Jurevičienė, D., Iljins, J. & Gaile-Sarkane, E. (2018). Factors influencing a bank's competitive ability: the case of Lithuania and Latvia. *Oeconomia Copernicana*, 9(1), 7-28.

[33] Simionescu M., Lazányi K., Sopková G., Dobeš K., & Balcerzak A. P. (2017). Determinants of Economic Growth in V4 Countries and Romania. *Journal of Competitiveness*, 9(1), 103-116.

[34] Smolović, J.C., Lipovina-Božović, M. & Vujošević, S. (2017). GARCH models in value at risk estimation: empirical evidence from the Montenegrin stock exchange. *Economic Research-Ekonomska Istraživanja*, 30(1), 477-498,

[35] Szumilo, N., Bienert, S., Łaszkiewicz, E., Pietrzak, M. B. & Balcerzak, A. P. (2018). The real alternative? A comparison of German real estate returns with bonds and stocks. *Journal of Property Investment & Finance*, 36(1), 19-31.

[36] Śliwiński, P., & Łobza, M. (2017). The impact of global risk on the performance of socially responsible and conventional stock indices. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(4), 657-674.

[37] Vukovic, D., Hanic, E. & Hanic, H. (2017). Financial integration in the European Union - the impact of the crisis on the bond market. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 12(2), 195-210.

[38] Zinecker, M., Balcerzak, A. P., Fałdziński, M., Meluzín, T. & Pietrzak, M. B. (2016). 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.* Vratna: Letra Interactive, 416-421.

[39] Zineker, M.; Łaszkiewicz, E., Meluzin, T., Pietrzak, M. B., & Balcerzak, A. P. (2017). Assessment of Changes in the Trend of Interdependences between the Capital Market of Germany and the Markets of Poland, the Czech Republic and Hungary. In J. Nešleha, T. Plíhal & K. Urbanovský (Eds.). *European Financial Systems 2017. Proceedings of the 14th International Scientific Conference Part 2***.** Brno: Masaryk University, 492-500.

[40] Zygmunt, A. (2017). Innovation activities of Polish firms. Multivariate analysis of the moderate innovator countries. *Oeconomia Copernicana*, 8(4), 505-521.

Data envelopment analysis models with penalties

Petra Zýková[1]

Abstract. The paper deals with advanced voting systems for ranking of candidates. The main aims of the system are to find a general winner and ranking of all candidates. Every voter gives ranking of the first *t*-candidates and also can give penalties to candidates who he/she surely does not want to vote for. There are advanced voting systems being used based on the application of data envelopment analysis (DEA) models. Original contribution of the paper consists in formulation of DEA/AR (assurance region) model with penalty and DEA/AR exclusion model with penalty. These models are derived from standard DEA/AR and DEA/AR exclusion models. The proposed models and they results are illustrated on Formula 1 World Constructors' Championship data set in season 2016 and the winner is determined.

JEL Classification: C44

AMS Classification: 90C05, 90C90

1 Introduction

Often, decision makers need to find a general winner based on the ranking of candidates in several voting rounds. This task may be solved by various voting systems with ranking of candidates. The problem consists in setting the weights (points) to particular ranks. Of course this setting influences final winner or complete ranking of candidates significantly. There is a space for proposal of models that specify the vector of weights. One of the possibilities is to apply data envelopment analysis (DEA) models. DEA models have been first developed by Charnes et al. [1] based on the concept introduced by Farrell [2]. Using DEA models, the optimal vector of weights for every candidate can be derived. This approach should make candidates confident in impartial and fair evaluation of voters' preferences. This use of DEA models was shown in Hashimoto [3]. In this way, the final ranking of candidates is determined. In many cases, the voters are sure that a candidate is unacceptable for them. Then they can assign him/her a negative weight (penalty). An original contribution of this paper is definition of two models with penalty.

The problem we are dealing with can be defined as follows: Let us assume a voting system with ranking of *n* candidates and *m* voters. Every voter gives ranking of the first *t*-candidates. This means, that he/she has not to assign ranking to all candidates but only to the top *t*-candidates $(t \le n)$. Voter must not give the same ranking to any two candidates. The aim is to find a general winner and general ranking of all candidates. The general ranking of candidates is obtained according to points, which *j*-th candidate gets from voters:

$$s_j = \sum_{r=1}^{t} u_r y_{rj} \quad j = 1,\ldots,n$$

where u_r, $r = 1,\ldots,t$ is the number of points gained for the r^{th} place, y_{rj}, $r = 1,\ldots,t$ and $j = 1,\ldots,n$ says how many times the j^{th} candidate was in the r^{th} place. We assume that the values y_{rj} are known. The general winner is the candidate with the highest value of s_j.

The paper is organized as follows. The next section presents definition of all models used in the study, i.e. DEA models, DEA/AR model, DEA/AR exclusion model, DEA/AR model with penalty and DEA/AR exclusion model with penalty. The following section contains numerical illustration on the real data set – determination of the winner of Formula 1 World Constructors' Championship 2016. Final section of the paper summarizes the results of all models.

2 DEA models

DEA models are a general tool for efficiency and performance evaluation of the set of homogenous DMUs that spend multiple (*w*) inputs and transform them into multiple (*t*) outputs. Measure of efficiency (efficiency score) of

[1] University of Economics Prague, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, petra.zykova@vse.cz.

this transformation is one of the main results of application of DEA models. Let us denote $\mathbf{Y} = \left(y_{rj}, r = 1,\ldots,t, j = 1,\ldots n\right)$ a positive matrix of outputs and $\mathbf{X} = \left(x_{kj}, k = 1,\ldots,w, j = 1,\ldots n\right)$ a positive matrix of inputs. The efficiency score of the unit under evaluation DMU_{j_0} is as follows:

$$\text{Maximize} \qquad U_{j_0} = \frac{\sum_{r=1}^{t} u_r y_{rj_0}}{\sum_{k=1}^{w} v_k x_{kj_0}}$$

$$\text{subject to} \qquad \frac{\sum_{r=1}^{t} u_r y_{rj}}{\sum_{k=1}^{w} v_k x_{kj}} \le 1, \quad j = 1,\ldots,n \qquad (1)$$

$$u_r \ge \varepsilon, \quad r = 1,\ldots,t$$

$$v_k \ge \varepsilon, \quad k = 1,\ldots,w,$$

where u_r is a positive weight of the r^{th} output, v_k is a positive weight of the k^{th} input, and ε is an infinitesimal constant. Model (1) is not linear in its objective function but may easily be transformed into a linear program. The linearized version of input-oriented model (often called CCR model) is as follows:

$$\text{Maximize} \qquad U_{j_0} = \sum_{r=1}^{t} u_r y_{rj_0},$$

$$\text{subject to} \qquad \sum_{k=1}^{w} v_k x_{kj_0} = 1, \qquad (2)$$

$$\sum_{r=1}^{t} u_r y_{rj} - \sum_{k=1}^{w} v_k x_{kj} \le 0, \quad j = 1,\ldots,n,$$

$$u_r \ge \varepsilon, \quad r = 1,\ldots,t,$$

$$v_k \ge \varepsilon, \quad k = 1,\ldots,w.$$

2.1 DEA without explicit inputs

In numerous applications a situation may occur that there are no inputs or there is one identical (unit) input. In such case a model without explicit inputs derived from previous models (1) and (2) can be formulated. This formulation is below:

$$\text{Maximize} \qquad U_{j_0} = \sum_{r=1}^{t} u_r y_{rj_0}$$

$$\text{subject to} \qquad \sum_{r=1}^{t} u_r y_{rj} \le 1, \quad j = 1,\ldots,n \qquad (3)$$

$$u_r \ge \varepsilon, \quad r = 1,\ldots,t.$$

An application of DEA models for the task we study avoids problems with determination of the vector of weights. This is the reason why the decision maker cannot manipulate the final ranking of candidates. The model is solved for every candidate and it is based on the assumption that it is fair used for every candidate a vector of weights which is the most favorable for him/her.

2.2 DEA model with assurance region (DEA/AR)

DEA/AR (assurance region) model (4) is the first model suitable for solving our task.

$$\text{Maximize} \qquad h_{j_0} = \sum_{r=1}^{t} u_r y_{rj_0}$$

subject to
$$\sum_{r=1}^{t} u_r y_{rj} \le 1,\ j = 1,\ldots,n \qquad (4)$$
$$u_r - u_{r+1} \ge \varepsilon,\ r = 1,\ldots,t-1$$
$$u_t \ge \varepsilon.$$

In comparison to model (3) this model contains last two set of constraints. They specify that the weight of a higher rank is greater than the weight of the immediately lower rank at least by ε and the weight of the last rank must be at least ε. The efficient unit(s) or candidate(s) obtains their efficiency scores h_{j_0} equal 1. This means that more than one candidate can be the winner which is not desirable. That is why, using of this model for our purposes is limited.

2.3 DEA/AR exclusion model

The previous model (4) does not guarantee obtaining a unique ranking of candidates. This means the results of the model can lead to the identification of more than one candidate in the first place. DEA/AR exclusion model (5) eliminates this drawback. Its formulation is as follows:

Maximize
$$g_{j_0} = \sum_{r=1}^{t} u_r y_{rj_0}$$

subject to
$$\sum_{r=1}^{t} u_r y_{rj} \le 1,\ j = 1,\ldots,n,\ j \ne j_0 \qquad (5)$$
$$u_r - u_{r+1} \ge \varepsilon,\ r = 1,\ldots,t-1$$
$$u_t \ge \varepsilon,$$
$$u_r - 2u_{r+1} + u_{r+2} \ge 0,\ r = 1,\ldots,t-2.$$

This model is a certain version of Andersen and Petersen super-efficiency model [1] that allows discrimination among originally efficient units. The candidate under evaluation is removed from the first set of constraints of model (5). In this way the units identified as efficient by model (4) have their efficiency score g_{j0} greater than 1 and so, they can be ranked easily.

3 Formulation of the models with penalties

In addition to the original formulation of the problem in the beginning of this paper, every voter can define negative candidates that are not acceptable for him/her. The aim is to find a general winner and a general ranking of all candidates. Let us assign the index (t+1) and the weight u_{t+1} to the negative ranking (penalty). The addition of penalty extends the DEA/AR and DEA/AR exclusion models (4) and (5) by new constraints.

There are two new constraints in models with penalty. The first constraint $u_{t+1} \le -\varepsilon$ guarantees that the negative weight cannot be equal to 0. The second one $\sum_{r=1}^{t+1} u_r y_{rj} \ge -1$ guarantees that assignment of one of the voters as negative does not automatically relegate the candidate to the last place.

3.1 DEA/AR with penalty

The first DEA model with penalty is DEA/AR (assurance region) model with penalty (6). This model is derived from model (4). Formulation of this model is the following:

Maximize
$$h^{-}_{j_0} = \sum_{r=1}^{t+1} u_r y_{rj_0},$$

subject to
$$\sum_{r=1}^{t+1} u_r y_{rj} \le 1,\ j = 1,\ldots,n, \qquad (6)$$

$$
\begin{aligned}
& u_r - u_{r+1} \geq \varepsilon\,,\ r = 1,\ldots,t-1\,,\\
& u_t \geq \varepsilon\,,\\
& \sum_{r=1}^{t+1} u_r y_{rj} \geq -1\,,\ j = 1,\ldots,n\,,\\
& u_{t+1} \leq -\varepsilon\,.
\end{aligned}
$$

3.2 DEA/AR exclusion model with penalty

The second DEA model with penalty is DEA/AR exclusion model with penalty derived from model (5). Its formulation follows:

$$
\begin{aligned}
\text{Maximize} \quad & g^{-}_{j_0} = \sum_{r=1}^{t+1} u_r y_{rj_0}\,,\\
\text{subject to} \quad & \sum_{r=1}^{t+1} u_r y_{rj} \leq 1\,,\ j = 1,\ldots,n\,,\ j \neq j_0\,, \qquad (7)\\
& u_r - u_{r+1} \geq \varepsilon\,,\ r = 1,\ldots,t-1\,,\\
& u_t \geq \varepsilon\,,\\
& \sum_{r=1}^{t+1} u_r y_{rj} \geq -1\,,\ j = 1,\ldots,n\,,\\
& u_{t+1} \leq -\varepsilon\,,\\
& u_r - 2u_{r+1} + u_{r+2} \geq 0\,,\ r = 1,\ldots,t-2\,.
\end{aligned}
$$

4 A numerical illustration

The application of the above mentioned models is illustrated on a data set of Formula 1 World Constructors' Championship results in season 2016. The results of Championship are presented in Table 1. Every team had two drivers. The column called technical fault will be considered as penalty in our models. The data set was obtained from the public website with Formula 1 results [4].

Team	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**technical fault**
Mercedes AMG Petronas	19	8	6	2	2	0	2	0	0	0	1
Red Bull Racing	2	8	6	8	4	2	4	2	0	0	2
Scuderia Ferrari	0	5	6	8	7	6	0	0	2	0	5
Sahara Force India	0	0	2	2	1	3	8	7	2	5	4
Williams Martini Racing	0	0	1	1	4	3	1	5	8	5	4
McLaren Honda Formula	0	0	0	0	2	2	4	1	4	4	6
Haas	0	0	0	0	1	1	1	1	0	1	8
Scuderia Toro Rosso	0	0	0	0	0	4	0	5	4	3	8
Renault Sport	0	0	0	0	0	0	1	0	0	2	8
Sauber	0	0	0	0	0	0	0	0	1	0	6
Manor Racing MRT	0	0	0	0	0	0	0	0	0	1	3

Table 1 Frequency of placement of teams and number of technical faults (source [5])

There are results of DEA models mentioned above in Table 2. Obtained values seem to be identical in models with and without penalty. It is caused by the given data set because the differences between teams are huge.

Team	DEA/AR	DEA/AR exclusion model	DEA/AR penalty	DEA/AR exclusion penalty
Mercedes AMG Petronas	1.0000	9.5000	1.0000	9.5000
Red Bull Racing	0.9231	0.9231	0.9231	0.9231
Scuderia Ferrari	0.8718	0.8718	0.8718	0.8718
Sahara Force India	0.7692	0.7692	0.7692	0.7692
Williams Martini Racing	0.7179	0.7179	0.7179	0.7179
McLaren Honda Formula	0.4359	0.4359	0.4359	0.4359
Haas	0.1282	0.1282	0.1282	0.1282
Scuderia Toro Rosso	0.4103	0.4103	0.4103	0.4103
Renault Sport	0.0769	0.0769	0.0769	0.0769
Sauber	0.0256	0.0256	0.0256	0.0256
Manor Racing MRT	0.0256	0.0256	0.0256	0.0256

Table 2 Efficiency scores given by models (4) to (7)

The highest efficiency score given by model (4) is 1.0000 for the team Mercedes AMG Petronas. Thus, only one team is efficient according to model (4). In is clear that it is not necessary to apply the super-efficiency model (5) in our case because we have just one efficient unit (team) but for illustration purposes, we present the results for model (5) in the second column of Table 2 as well. The highest efficiency (super-efficiency) score given by model (5) is 9.5000 for the team Mercedes AMG Petronas. As presented in Table 2, models with penalty (6) and (7) lead to the same efficiency and super-efficiency scores for this team.

The final ranking of teams according to all models mentioned above is presented in Table 3. All applied models allow complete ranking of all teams in our case. The placement on 10^{th} and 11^{th} place is given by the next numbers behind the decimal point. The data show that all rankings are similar. The first team and the winner of World Constructors' Championship is Mercedes AMG Petronas, second is the Red Bull Racing team and the third is the Scuderia Ferrari. The ranking of the teams by models with and without penalty are same in our case. Of course, this conclusion does not hold in general.

Team	DEA/AR	DEA/AR exclusion model	DEA/AR penalty	DEA/AR exclusion penalty
Mercedes AMG Petronas	1	1	1	1
Red Bull Racing	2	2	2	2
Scuderia Ferrari	3	3	3	3
Sahara Force India	4	4	4	4
Williams Martini Racing	5	5	5	5
McLaren Honda Formula	6	6	6	6
Haas	8	8	8	8
Scuderia Toro Rosso	7	7	7	7
Renault Sport	9	9	9	9
Sauber	10	10	11	11
Manor Racing MRT	11	11	10	10

Table 3 Ranking of candidates by models (4) to (7)

5 Conclusions

The determination of a general winner and complete ranking of candidates is an interesting task whose solution is significantly influenced by the vector of weights assigned to the candidates. This paper deals with an application of data envelopment analysis models that return a unique optimal vector of weights for particular candidates. This unique vector ensures that any candidate should not be disadvantaged. DEA/AR and DEA/AR exclusion models

are the first DEA models used in this paper. These two models work only with positive rankings by voters. In contrast, DEA/AR model with penalty and DEA/AR exclusion model with penalty work with the rankings given by voters and also with the negative points from voters.

All models were practically demonstrated on determination of a general winner and complete ranking of 11 teams in Formula 1 World Constructors' Championship in season 2016. The demonstration is a rather illustrative case and not a serious study. The results are similar with and without penalty which is given by the data set used. This conclusion does not hold true in general. As expected the winner according to all models used in this paper is the team Mercedes AMG Petronas. The numerical experiment were realized using original procedures written in LINGO modeling language. Voting systems are an interesting subject for further research as the application of DEA models in this field is not very frequent. In our future research we plan to extend the presented idea and to apply it in more extensive a serious case studies.

Acknowledgements

This work was supported by the Internal Grant Agency of the Faculty of Informatics and Statistics, University of Economics, Prague, project F4/20/2018 (Data envelopment analysis models in economic decision making).

References

[1] Andersen, P. and Petersen, N. (1993). A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science*, 39(10), pp. 1261-1264.
[2] Charnes, A., Cooper, W. and Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), pp. 429-444.
[3] Farrell, M. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), p. 253.
[4] Hashimoto, A. (1997) A ranked voting system using a DEA/AR exclusion model: A note. *European Journal of Operational Research*, 97(3): 600-604.
[5] http://www.vysledky.cz/automobily/formule-1/

Assessment of Cluster Benefits to Increase Financial Performance of Companies: Malmquist Index Approach

Miroslav Žižka[1]

Abstract. Cluster organizations are one of the tools to support regional innovation performance. The creation and development of clusters is supported by the EU Structural Funds. The issue of efficient use of public resources is therefore very important. The aim of the article is to determine whether member companies of the cluster organization in the technical textiles sector achieve better financial performance in the time series 2009-15 than companies operating in the same industry, the same region; however they are not members of a cluster organization. The first sample includes 16 members of the cluster organization. The second sample consists of 45 companies from the Northeast Cohesion Region doing business in the technical textiles sector. Data envelopment analysis with three inputs and one output was used to evaluate the efficiency. Inputs are employee numbers, total assets and capital employed, and output is economic value added (EVA). Since EVA has both positive and negative values, a variant of the radial VRM model was used. For both samples and periods, the Malmquist index values were calculated. The Malmquist index breaks down the overall change in factor performance into two components - to technical efficiency change and to technological change. The research has shown a more significant improvement in financial performance at companies in cluster compared to other industry sectors. Improvement in performance was driven mainly by technological change, it means by shifting the efficient frontier.

Keywords: Malmquist index, VRM model, cluster of technical textiles, economic value added, financial performance.

JEL Classification: C61, L25, L67
AMS Classification: 90B90, 90C90

1 Introduction

The paper deals with the influence of the cluster on the growth of financial performance of member companies. The research builds on the previous article [12], which evaluated the overall change in factor productivity and its components in CLUTEX clusters in 2009-14 using the Malmquist index. Several major changes were made in further research. The time series was extended by another period (2015) and a second research sample was added. In this second sample there are companies from the same industrial sector as the CLUTEX cluster which operate in the same region as the cluster (the three regions - Liberec, Hradec Králové and Pardubice, which form the cohesion region Northeast), but are not members of a cluster organization. Another change applies to the model used. The output (economic value added) is both positive and negative at individual DMUs, which is contrary to the requirements of traditional DEA models. In the previous research, negative output values were modified by the ratio index construction and by adding a large positive value. In the new research, the issue of negative values is treated with the VRM radial model variant.

The aim of the research is to find out whether CLUTEX cluster companies had a more significant performance improvement than other companies that are not members of this cluster. A partial goal is to identify components that contributed significantly to the expected growth in performance.

2 Malmquist index and Data Envelopment Analysis

In 1953, Malmquist proposed a quantitative index originally for the use in consumer analysis. Later, the scope of its application was expanded to production, where it can be used to generate output quantity indices, input quantities, or productivity indices as shares of input or output distance functions [5]. The main advantage of the Malmquist *MI* indexes is the possibility of decomposition of the change of productivity into the change of the technical efficiency and the technological change between the two periods, see the relation (1). The Malmquist index works with panel data. The efficiency change *EFFCH* expresses a catch-up effect, it means the extent to which the DMU approaches the most efficient production frontier. The change in efficiency reflects the

[1] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec 1, miroslav.zizka@tul.cz.

organization's internal efforts to improve [8]. Technological change *TECH* is characterized by an efficiency frontier shift as a result of industry innovation. The relationships for calculating both types of changes for the input oriented Malmquist index are given in equations (4) and (5). The *MI*, *EFFCH* and *TECH* values that are greater than one mean productivity growth, efficiency gains, and technological progress. On the contrary, values lower than one indicate deterioration.

Both components can be further decomposed. The efficiency change *EFFCH* to the product of a change in pure technical efficiency *PECH* and the scale efficiency change *SECH*. The first component expresses the unit's ability to improve technical efficiency between periods *t* and *t + 1* under variable returns to scale conditions. The second *SECH* component measures the scale efficiency change between periods. If *SECH* is greater than one, the outputs in *t +1* are closer to the technical optimum than in the previous period. The unit achieves a technologically optimal production range when it is in constant returns to scale (CRS) conditions [10]. Similarly, the *TECH* technology change can be broken down into the product of three components. The first two components *OBITECH* and *IBITECH* measure the distance between the effective frontiers between periods, but in relation to the change in the input/output proportions. *OBITECH* takes into account the vector of inputs from the period of *t + 1* and the vector of outputs from both periods. *IBITECH* takes into account the vector of the outputs of period *t* and the input vector from both periods [5]. Both indicators compare the extent of technological change with the ray in the period of *t + 1* to the extent of technological change with the ray in period *t*. If the bias indexes are equal to one, the changes in the proportions of inputs and outputs do not affect the change in productivity. The change in production function is then referred to as Hicks-neutral technological change [5]. The third component, technical change magnitude *TECHM*, measures the relative distance of effective frontiers between periods under the conditions of input and output neutrality. Formulas (1) to (10) were taken from [6].

$$MI_q(x^{t+1}, y^{t+1}, x^t, y^t) = EFFCH_q TECH_q \tag{1}$$

$$EFFCH_q = PECH_q SECH_q \tag{2}$$

$$TECH_q = TECHM_q OBITECH_q IBITECH_q \tag{3}$$

$$EFFCH_q = \frac{D_q^{t+1}(x^{t+1}, y^{t+1})}{D_q^t(x^t, y^t)} \tag{4}$$

$$TECH_q = \sqrt{\frac{D_q^t(x^t, y^t)}{D_q^{t+1}(x^t, y^t)} \frac{D_q^t(x^{t+1}, y^{t+1})}{D_q^{t+1}(x^{t+1}, y^{t+1})}} \tag{5}$$

$$PECH_q = \frac{D_{qv}^{t+1}(x^{t+1}, y^{t+1})}{D_{qv}^t(x^t, y^t)} \tag{6}$$

$$SECH_q = \sqrt{\frac{D_{qv}^t(x^{t+1}, y^{t+1}) / D_{qc}^t(x^{t+1}, y^{t+1})}{D_{qv}^t(x^t, y^t) / D_{qc}^t(x^t, y^t)} \frac{D_{qv}^{t+1}(x^{t+1}, y^{t+1}) / D_{qc}^{t+1}(x^{t+1}, y^{t+1})}{D_{qv}^{t+1}(x^t, y^t) / D_{qc}^{t+1}(x^t, y^t)}} \tag{7}$$

$$OBITECH_q = \sqrt{\frac{D_q^t(x^{t+1}, y^{t+1})}{D_q^{t+1}(x^{t+1}, y^{t+1})} \frac{D_q^{t+1}(x^{t+1}, y^t)}{D_q^t(x^{t+1}, y^t)}} \tag{8}$$

$$IBITECH_q = \sqrt{\frac{D_q^t(x^{t+1}, y^t)}{D_q^t(x^{t+1}, y^t)} \frac{D_q^{t+1}(x^t, y^t)}{D_q^t(x^t, y^t)}} \tag{9}$$

$$TECHM_q = \frac{D_q^t(x^t, y^t)}{D_q^{t+1}(x^t, y^t)} \tag{10}$$

$D^t(x^t, y^t)$ ... the efficiency score of period *t* referring to the technology of period *t*,
$D^{t+1}(x^{t+1}, y^{t+1})$ ... the efficiency score of period *t* + 1 referring to the technology of period *t* + 1,
$D^{t+1}(x^t, y^t)$ ... the efficiency score of period *t* referring to the technology of period *t* + 1,
$D^t(x^{t+1}, y^{t+1})$ ... the efficiency score of period *t* + 1 referring to the technology of period *t*,
$D^1(x^t, y^{t+1})$ ... the efficiency score of the virtual DMU $[x^t, y^{t+1}]$ referring to the technology of period *t*,
$D^{t+1}(x^t, y^{t+1})$ ... the efficiency score of the virtual DMU $[x^t, y^{t+1}]$ referring to the technology of period *t* + 1,
c ... constant returns to scale (CRS),
v ... variable returns to scale (VRS).

Data Envelopment Analysis was used to calculate the values of the distance functions. Before the formulation of the model, a problem with negative economic value added had to be solved. EVA gains both

positive and negative values. This problem can be solved in several ways. If all output values were negative, they could be transformed into positive inputs [11]. However, such a transformation is not possible in this case because EVA gains negative values for some DMUs and positive values for other units. Another possible approach is to add a large positive constant to all the values of the variable. Such an approach was applied in previous research, [12] where EVA was transformed into a ratio indicator to reduce the size of a constant. Nevertheless, such an approach leads to the efficient frontier shifts and causes interpretation problems.

For these reasons, a model with a variant radial measure (VRM) was designed [7] and is suitable for the positive and negative values of some of the variables. The VRM model input-oriented equations (11) are presented below. Modification of the model lies in using an absolute value of inputs (outputs) instead of their actual values. The difference $1 - \beta$ expresses the efficiency assessed by the DMU. Thus, β is a measure of inefficiency and indicates the degree of improvement required to reach the frontier when applying a proportional reduction of inputs [7].

$$\max \beta$$

$$\begin{aligned} &\boldsymbol{X}\lambda + \beta\left|\boldsymbol{x}_q\right| \le \boldsymbol{x}_q \\ &\boldsymbol{Y}\lambda \ge \boldsymbol{y}_q \\ &\sum_j \lambda_j = 1 \\ &\lambda \ge 0 \end{aligned} \tag{11}$$

3 Data and methodology

To assess the financial performance of textile companies, it was necessary to obtain data from the balance sheet and profit and loss statement for the period 2009-15. In addition, data was also gained on the number of employees for the same period. All data comes from the MagnusWeb [1] database. The beginning of the time series was chosen taking into account the CLUTEX cluster's launch in 2007, since the positive effect of the cluster on member organizations has a certain time lag. The end of the time series in 2015 is given by the availability of financial statements. At the beginning of 2018, there were not yet available reports for the year 2016 for a large number of companies. It should be noted that although the Accounting Act imposes the obligation to disclose the financial statements on companies listed in the public register, a number of enterprises do not fulfill this obligation. According to CRIF's analysis, 69% of companies that have this statutory obligation did not publish their financial statements in 2014 [2].

The research was carried out in the following steps:

1. **Establishing a list of companies in both research sets** - the first research set is made up of member companies that form the core of CLUTEX cluster organization. The CLUTEX cluster currently has 28 members (February 2018). Compared with the previous year, therefore, the number of members decreased by five. The activities of the member organizations are very heterogeneous - from the production of textiles and clothing, through business to research and education. Business performance can be compared only at homogeneous production units. On the basis of an analysis of cluster member's industry focus, the NACE 13200, 13900 and 14100 sectors were identified as the core of the cluster. There are 19 such companies. However, the financial statements in the complete time series 2009-15 were only obtained for 16 companies which represent the first sample. The second research set involves companies that operate in the same areas of the Liberec, Hradec Králové and Pardubice Regions and are not members of CLUTEX cluster. Each company may be part of only one sample. There are 168 companies of this kind. The financial statements in the complete time series 2009-15 were obtained from 45 companies that form the second research sample. The lower success rate of obtaining the financial statements in the second set is due to the fact that the latter sample was represented to a larger extent by smaller companies that are not obliged to publish their financial statements.
2. **Input and output specification** - basic factors of production were selected as the inputs of the model – the number of employees, the total assets and the long-term capital of the company. The output is the economic value added (EVA) that represents net operating profit after deducting capital costs. EVA was established according to the methodology of the Ministry of Industry and Trade [9]. The basic formula is given in the equation (12). The process of determining the individual indicators is described in the MPO methodology. EVA can have both positive and negative values. In the case of a positive EVA, the company generates value for its owners. If EVA is negative, then the value of the company decreases. To eliminate the impact of the price change, financial data at constant prices in 2009 was used. For the conversion of current prices, the industrial producer price index was used in the textile, clothing and leather industry, published by the CZSO [4].

$$EVA = (ROE - r_e).E \tag{12}$$

ROE ... return on equity,
r_e ... alternative cost of equity,
E ... equity.

3. **Determination of the Malmquist index and its decomposition** – for each DMU and period, the values of the distance functions were determined through DEA and then the individual components of the Malmquist index according to relations (2) to (10). Finally, the value of the Malmquist index was calculated by relation (1). For each period and research sample (CLUTEX, other textile companies), the geometric mean of Malmquist index and its components were then calculated. The MaxDEA 7 Ultra [6] software was used to calculate the values of the distance functions.
4. **Comparison of the Malmquist Index and its components among the research samples** – non-parametric Kolmogorov-Smirnov test was used to compare performance differences between clustered CLUTEX companies and non-clustered companies. The test compares the distribution of two samples (CLUTEX cluster companies and other companies). Test statistics *DN* indicates the maximum distance between the cumulative distributions of the two samples. The non-parametric test was chosen because the values of the individual variables did not have normal distribution, see Table 3. Shapiro-Wilk test was used to test normality. Statistical testing was performed using STATGRAPHICS Centurion XVII.

4 Research results

The Malmquist index values for both of the research samples is shown in Table 1 and 2. It follows from both tables that the index value fluctuated considerably within individual years. In the sample of CLUTEX cluster member companies, there is recorded a decline in 2011-13 following a growth in 2009-11. Thereafter, the financial performance increased in 2013-14 and a relatively surprisingly declined (due to economic growth, see the last column in Table 1) over the years 2014-15. In the second sample of other textile companies, the value of the index declined in 2009-10, followed by a rise in 2010-12 and again by a decline in the next period. Further, the sample was similar to the former one, with the value of the index increasing over the period 2013-14 and also surprisingly declining in the last reference period.

It can be assumed that the general economic situation in the Czech Republic influenced the development of the performance of textile companies. In the years 2010 to 2011, the Czech economy showed GDP growth. In the next two years, the Czech economy was in recession and since 2014 it has been showing economic growth again. This partly corresponds to the Malmquist index values (the *MI* dependence on economic growth cannot be verified for a short time series, the Spearman rank correlation coefficient is 0.43, but the p-value is 0.24). It is clear from Table 1 that in the years of economic growth, the Malmquist index values in clustered companies were bigger than one, thus increased their performance. By contrast, in 2012-13, the value of the Malmquist index was less than one, which corresponds to the recession period. Only the last period of 2014-15, when companies reported a decrease in financial performance, does not follow this trend, although the Czech economy was growing. The reasons for this fluctuation are not clear. It could be caused by a short-term fluctuation in the industry. If developments in the preceding period 2015/14 indicated some structural problems in the industry, it would be necessary to verify them on the dates of the following period 2016/15.

Year	MI	EFFCH	PECH	SECH	TECH	OBITECH	IBITECH	TECHM	Ec. growth
2010/2009	3.4100	2.3722	6.2784	0.3778	1.4375	1.1354	1.0345	1.2238	1.0227
2011/2010	2.0096	1.4416	2.7323	0.5276	1.3940	1.0513	1.0007	1.3250	1.0178
2012/2011	0.7261	0.6682	0.4458	1.4989	1.0866	0.9483	1.0032	1.1422	0.9920
2013/2012	0.6383	0.6383	0.3553	1.7964	1.0000	1.0000	1.0000	1.0000	0.9952
2014/2013	1.1475	0.7067	1.0805	0.6541	1.6238	1.2606	1.0000	1.2882	1.0272
2015/2014	0.9802	0.9802	1.0557	0.9285	1.0000	1.0000	1.0000	1.0000	1.0531
GEOMEAN	**1.2364**	**1.0017**	**1.2075**	**0.8296**	**1.2343**	**1.0610**	**1.0063**	**1.1560**	**1.0178**

Table 1 Malmquist index summary of annual means in the CLUTEX cluster

A rather different development was identified in the second sample of other textile companies. In the first monitored period the performance decreased, which may have been a consequence of the previous economic crisis (in 2009 the Czech economy was in recession - it showed a decline of 4.8% [3]). Improvements in the performance of textile companies in the second set continued until 2012, when the first set of cluster companies experienced a decline in performance. In 2014-13, there was a much stronger performance recovery than for

companies in the first sample. In the latter period, however, the second sample also showed a decline in financial performance.

Year	MI	EFFCH	PECH	SECH	TECH	OBITECH	IBITECH	TECHM	Ec. growth
2010/2009	0.9718	0.9483	0.8913	1.0641	1.0247	1.0046	1.0044	1.0155	1.0227
2011/2010	1.2829	1.2270	1.7362	0.7067	1.0455	1.0211	1.0018	1.0221	1.0178
2012/2011	1.0568	0.9863	0.8792	1.1218	1.0714	0.9967	1.0069	1.0676	0.9920
2013/2012	0.7360	0.7326	0.7237	1.0124	1.0046	1.0008	0.9995	1.0043	0.9952
2014/2013	1.7799	1.7702	3.9958	0.4430	1.0054	1.0008	1.0003	1.0043	1.0272
2015/2014	0.8805	0.8805	0.6697	1.3147	1.0000	1.0000	1.0000	1.0000	1.0531
GEOMEAN	**1.0722**	**1.0461**	**1.1752**	**0.8901**	**1.0250**	**1.0040**	**1.0021**	**1.0187**	**1.0178**

Table 2 Malmquist index summary of annual means for other textile companies

It can be stated that the financial performance of clustered companies increased more dramatically than in the sample of non-clustered companies. In 2009-15, company performance in a cluster organization increased by an average of 24% per year, while in other companies only by about 7% per year. Tables 1 and 2 also show that the performance of cluster companies was driven mainly by technological progress (innovations). Technological change was mainly due to a shift of magnitude, 16% for cluster companies and nearly 2% for the second sample of companies per year. Input and output bias indexes contributed only slightly to changing performance (index values are around one, with the exception of two periods for the OBITECH index for clustered companies). The situation is therefore close to the conditions of Hicks neutrality. CLUTEX cluster companies at the efficient frontier increased their performance by 23% per year, while in the second sample only by 2.5% per year. The internal relative efficiency of the clustered companies did not practically change. On the contrary, in the second sample of other companies, the main source of performance growth was the improvement of internal efficiency, on average by about 5% per year. For both samples, it applies that the improvement in internal efficiency was due to the growth of pure technical efficiency. The scale efficiency in both samples decreased, on average, by 17% per year for clustered companies, 11% per year for non-clustered companies.

Since the Shapiro-Wilk test showed (Table 3) that the indicators did not have a normal distribution, the non-parametric Kolmororov-Smirnov test was used to assess the differences. Based on the data in table 4, it can be concluded that there are significant differences in the change of overall performance, efficiency change and frontier shift between the two samples for the period of 2009-15 at alpha significance level of 5%. Only the difference between the scale efficiency change is not significant over the same period.

CLUTEX			Other textile companies		
Indicator	**Shapiro-Wilk test**	**P-Value**	**Indicator**	**Shapiro-Wilk test**	**P-Value**
MI	0.1726	<0.0001	MI	0.08138	<0.0001
EFFCH	0.1559	<0.0001	EFFCH	0.0800	<0.0001
PECH	0.1257	<0.0001	PECH	0.06394	<0.0001
SECH	0.1700	<0.0001	SECH	0.09255	<0.0001
TECH	0.2608	<0.0001	TECH	0.2694	<0.0001
TECHM	0.2261	<0.0001	TECHM	0.1519	<0.0001

Table 3 Shapiro-Wilk test of normality

Year	MI	EFFCH	PECH	SECH	TECH	TECHM
2009-2015	0.1694	0.2153	0.2042	0.1241	0.8296	0.5926
	0.0343	**0.0028**	**0.0055**	0.2261	**<0.0001**	**<0.0001**

Table 4 Kolmogorov-Smirnov Test, DN statistic (P-Values) – CLUTEX cluster vs. other companies

5 Conclusion

Based on the above research, it can be stated that a stronger growth in financial performance for CLUTEX cluster companies was proved compared to the other textile companies. Improving the financial performance of clustered companies was mainly due to a shift in efficient frontier, it means by technological change and innovation. Companies in the CLUTEX cluster also responded more quickly to the effects of the 2009 recession. Other non-member companies in the textile sector showed improved financial performance with a one-year delay compared to cluster companies. For non-clustered companies, the main source of financial performance growth was the improvement of relative technical efficiency.

This has confirmed the presumption of the positive influence of cluster organizations on innovation and financial performance of companies as formulated in previous research [12]. This assumption was confirmed by

empirical comparison of two samples of companies from the same industry (textile industry) operating in the same region (NUTS 2 Northeast). A rather surprising finding is the decline in financial performance in the last monitored period of 2014-15 for both company samples, despite the continuing economic growth of the Czech economy. In both samples of companies, this decline was attributable to the deterioration of technical efficiency, with the simultaneous stagnation of the technology shift. Another finding is probably the exhausted potential of economies of scale, especially in clustered companies.

Some differences from previous research are also worth highlighting. Previous research found the improvement in the financial performance of clustered companies driven by technological change, but also reduced relative technical efficiency. It should be noted that the samples of companies are not identical. In the meantime, the number of CLUTEX cluster member companies decreased by five organizations. This apparently had a negative effect on the scale efficiency, which decreased compared to the previous research.

Further research will focus on extending the time series to another year, tracking technological change in the textile sector (recent data suggest possible exhaustion of innovation potential) and expanding research to other sectors.

Acknowledgements

Supported by the grant No. GA18-01144S „An empirical study of the existence of clusters and their effect on the performance of member enterprises“ of the Czech Science Foundation.

References

[1] Bisnode: *Magnusweb: Komplexní informace o firmách v ČR a SR* [online]. Bisnode ČR, Prague, 2018 [cit. 2018-02-14]. Available at: https://magnusweb.bisnode.cz.

[2] CRIF: *Analýza: firmy, které nezveřejňují účetní závěrky, končí téměř 6x častěji v exekuci* [online]. Czech Credit Bureau, Prague, 2016-08-08 [cit. 2018-22-02]. Available at: http://www.informaceofirmach.cz/analyza-firmy-ktere-nezverejnuji-ucetni-zaverky-konci-temer-6x-castej-i-v-exekuci.

[3] CZSO: *Statistical Yearbook of the Czech Republic – 2017* [online]. Czech Statistical Office, Prague, 2017 [cit. 2018-03-17]. Available at: https://www.czso.cz/csu/czso/statistical-yearbook-of-the-czech-republic.

[4] CZSO: *Veřejná databáze – Indexy cen průmyslových výrobců* [online]. Czech Statistical Office, Prague, 2018 [cit. 2018-02-14]. Available at: https://vdb.czso.cz.

[5] Färe, R., Grifell-Tatjé, E., Grosskopf, S., and Knox Lovell, C.A.: Biased Technical Change and the Malmquist Productivity Index. *The Scandinavian Journal of Economics* **99** (1997), 119–127.

[6] Cheng, D.: *Data Envelopment Analysis: Methods and MaxDEA Software.* Intellectual Property Publishing House, Beijing, 2014.

[7] 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 Operational Research* **225** (2013) 100–105.

[8] Li, Z., Crook, J., and Andreeva, G.: Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems With Applications* **80** (2017), 94–106.

[9] MPO: *Finanční analýza podnikové sféry za rok 2015.* Ministry of Industry and Trade, Prague, 2017.

[10] Pantzios, Ch.J., Karagiannis, G., and Tzouvelekas, V.: Parametric decomposition of the input-oriented Malmquist productivity index: with an application to Greek aquaculture. *Journal of Productivity Analysis* **36** (2011), 21–31.

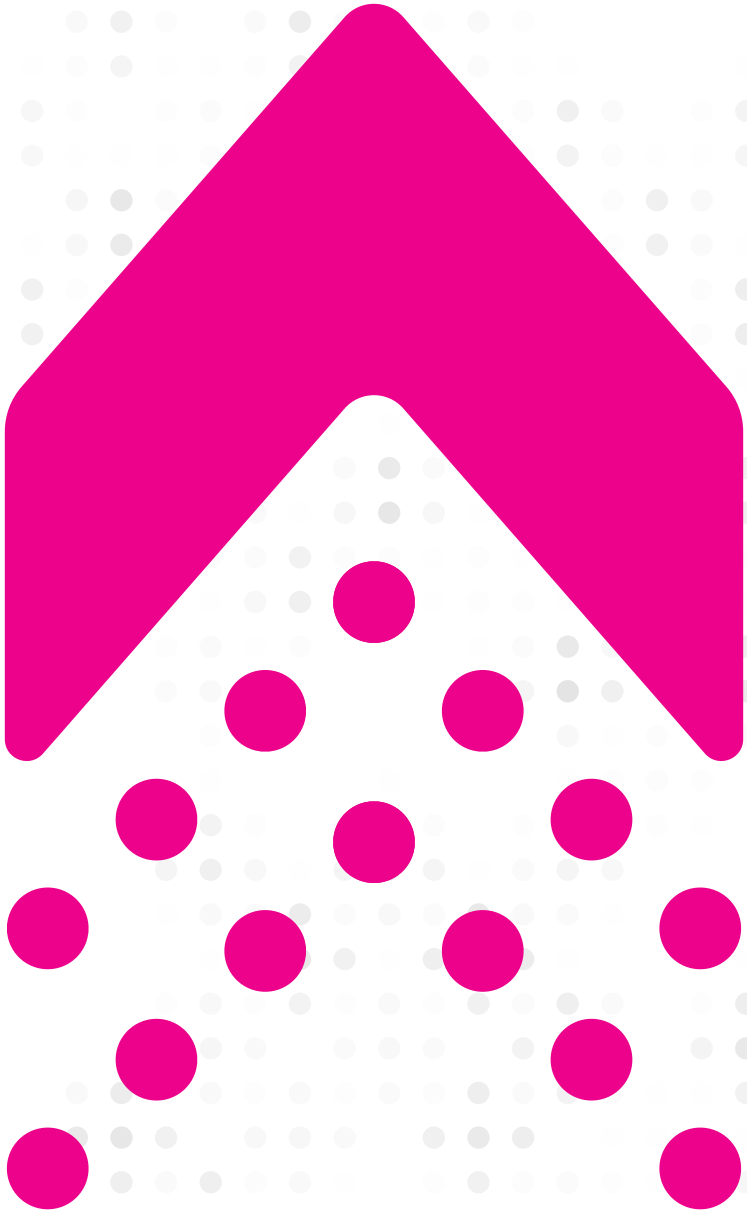
[11] Pastor J.T., Ruiz J.L. (2007) Variables With Negative Values In Dea. In: *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis* (Zhu, J., Cook, W.D., eds.). Springer, Boston, MA, 2007.

[12] Žižka, M.: Use of Malmquist Index in Evaluating Performance of Companies in Cluster. In: *35th International Conference Mathematical Methods In Economics, Conference Proceedings* (Žižka, M., ed.). Gaudeamus, Hradec Králové, 2017, 878–883.

List of Authors

VŠE FACULTY OF MANAGEMENT JINDŘICHŮV HRADEC

ČSOV



ISBN 978-80-7378-372-3

9 788073 783723